Maschinen als kollaborative Gesprächspartner.
Nutzer- und situationsorientierte Gestaltung automotiver Sprachdialogsysteme

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Linn Hackenberg
aus Altenburg

1. Referent: Prof. Dr. M. Vollrath
2. Referent: Prof. Dr. J. F. Krems
eingereicht am: 24.10.2012
mündliche Prüfung (Disputation) am: 18.12.2012
Druckjahr 2013

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Hackenberg, L. & Vollrath, M.: Overcoming Potential Barriers in Human-Machine-Communication. tubs.CITY Symposium 2010, Braunschweig (2010).

Hackenberg, L.: Collaborative Human-Machine Communication. (Vortrag & Poster) 13. IFIP TC13 Conference on Human-Computer Interaction, Lissabon (2011).

Hackenberg, L., Vollrath, M. & Oel, P.: Nutzerorientierte Feedbackgestaltung von Sprachdialogsystemen. (Vortrag & Paper) 6. VDI-Tagung: Fahrer im 21. Jahrhundert, Braunschweig (2011).

Hackenberg, L. & Neugebauer, M.: Investigating linguistic alignment in the domain of in-vehicle speech dialog systems. 54. Tagung experimentell arbeitender Psychologen, Mannheim (2012).

„Es war einmal ein armes frommes Mädchen, das lebte mit seiner Mutter allein, und sie hatten nichts mehr zu essen. Da ging das Kind hinaus in den Wald, und da begegnete ihm eine alte Frau, die wusste seinen Jammer schon und schenkte ihm ein Töpfchen, zu dem sollt es sagen »Töpfchen, koche,« so kochte es guten süßen Hirsebrei, und wenn es sagte »Töpfchen, steh,« so hörte es wieder auf zu kochen. Das Mädchen brachte den Topf seiner Mutter heim, und nun waren sie ihrer Armut und ihres Hungers ledig und aßen süßen Brei, so oft sie wollten. Auf eine Zeit war das Mädchen ausgegangen, da sprach die Mutter »Töpfchen, koche,« da kocht es, und sie isst sich satt; nun will sie, dass das Töpfchen wieder aufhören soll, aber sie weiß das Wort nicht. Also kocht es fort, und der Brei steigt über den Rand hinaus und kocht immerzu, die Küche und das ganze Haus voll, und das zweite Haus und dann die Straße, als wollts die ganze Welt satt machen, und ist die größte Not, und kein Mensch weiß sich da zu helfen. Endlich, wie nur noch ein einziges Haus übrig ist, da kommt das Kind heim, und spricht nur »Töpfchen, steh,« da steht es und hört auf zu kochen; und wer wieder in die Stadt wollte, der musste sich durchessen.“
(Gebrüder Grimm, 1812/1815, Stelle 103)

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als Doktorandin in der Konzernforschung der Volkswagen AG von 2009 bis 2012. Sie dokumentiert einen Teil des Erkenntnisgewinns, den ich während dieser Zeit erfahren durfte, und an dem viele Personen einen Anteil hatten, denen ich an dieser Stelle danken möchte.

Zunächst geht mein Dank an Prof. Dr. Mark Vollrath für seine immerwährende Gesprächsbereitschaft, die stets konstruktiven und motivierenden Beiträge und den herzlichen Umgang. Darüber hinaus danke ich Prof. Dr. Josef F. Krems für seine freundliche Bereitschaft, das Ko-Referat zu übernehmen.

Ebenfalls möchte ich Dr. Moritz Neugebauer ganz herzlich für das geduldige und ausdauernde Korrekturlesen der Arbeit danken. Als Vorbild und Freund konnte er durch zahlreiche kreative Anregungen und wissenschaftstheoretische Auseinandersetzungen zur Entstehung der Arbeit beitragen.

Ferner danke ich den Kollegen, die mir bei der Volkswagen AG inhaltlich und organisatorisch unterstützend zur Seite gestanden haben, speziell Dr. Peter Oel, Dr. Helge Neuner und Dr. Miklós Kíss, die mir stets Vertrauen entgegengebracht und die Entstehung der Arbeit ermöglicht haben. Darüber hinaus gilt mein Dank Gordon Seitz für die umfangreiche technische Unterstützung, Andreas Galla für die Hilfe bei den Simulatorstudien, Ina Petermann-Stock für zahlreiche Kaffeepausen im Doktorandenraum sowie den Kollegen des Probandenpools für die organisatorische Hilfe bei den Studien. Auch danken möchte ich den von mir betreuten Praktikanten, Diplomanden und Masterstudentinnen Axel Knobloch, Sonja Giesemann, Anke-Lisa Paul und Lutz Kolburg für ihre Unterstützung sowie die umfangreiche Hilfe bei der Durchführung der Studien.

Worte des Dankes reichen kaum aus, um der Hilfe und Unterstützung meiner Eltern gerecht zu werden, die immer bedingungslos hinter mir gestanden und nie an meinem Vorhaben gezweifelt haben.

Ebenfalls ungenügend bleiben alle Dankesworte, die versuchen, den Anteil von Benjamin Saß und Helena Katzmann an dieser Arbeit zu würdigen. Für mich sind sie der Hafen.

Disclaimer

Veröffentlichungen über den Inhalt der Arbeit sind nur mit schriftlicher Genehmigung der Volkswagen AG zugelassen.

Die Ergebnisse, Meinungen und Schlüsse dieser Dissertation sind nicht notwendigerweise die der Volkswagen AG.

Inhaltsverzeichnis

1 Einleitung

Moderne Infotainmentsysteme in Fahrzeugen zeichnen sich durch einen hohen Grad an Funktionsvielfalt und Komplexität aus. Der Großteil der Zusatzfunktionen in diesem Bereich dient ausschließlich der Unterhaltung und nicht dem unmittelbaren Zweck der Fahrsituation (Hamerich, 2009). Die Entwicklung von Mensch-Maschine-Schnittstellen, die dem Fahrer die Bedienung dieser Infotainmentfunktionen ermöglichen, aber zeitgleich die Ablenkung auf ein Minimum reduzieren, ist daher ein logischer Schritt zur Gewährleistung der Fahrersicherheit bei stetiger Erhöhung des Anspruchs an den Fahrkomfort.

Um die Anzahl der Blickabwendungen in Zukunft weiter zu reduzieren, liegt insbesondere im Einsatz von Sprachdialogsystemen (SDS) noch großes Potential. In verschiedenen Studien konnte belegt werden, dass der Einsatz von SDS die Ablenkungspotentiale klassischer grafisch-haptischer Bedienkonzepte im Fahrtkontext reduzieren kann (vgl. z. B. Tsimhoni, Smith & Green, 2004; Maciej & Vollrath, 2009). Neben diesem Sicherheitsaspekt gilt Sprache als eine der natürlichsten und intuitivsten Interaktionsmodalitäten und lässt das direkte Ansprechen von Funktionen ohne Beachtung von Menü-Hierarchien zu. Sprache ermöglicht zudem unscharfe Such-Anfragen und einen effizienten Zugriff auf komplexe Datenmengen. Aufgrund dieser umfangreichen Vorteile etablierte sich die Sprachbedienung im Fahrzeug und wird heute von nahezu allen Automobilherstellern angeboten. Die Unternehmensberatung Frost und Sullivan (2012) schätzt, dass bereits im Jahr 2017 rund 22.9 Millionen Autos in Europa mit einer sprachbasierten Nutzerschnittstelle ausgestattet sein werden.

Trotz der steigenden Verfügbarkeit und stetigen technologischen Weiterentwicklungen nutzen Menschen SDS im Fahrzeug weniger, als es ihr Potential vermuten lässt. Obwohl viele Fahrer den innovativen Charakter von Sprache als Bedienschnittstelle erkannt haben und motiviert sind davon Gebrauch zu machen, werden sie oftmals bei der Interaktion mit einem Sprachdialogsystems enttäuscht. Der Dialog wird als schwierig und anstrengend empfunden und die Interaktion bald wieder auf andere Modalitäten verlagert. Die geringe nutzerseitige Akzeptanz und Zufriedenheit lassen neue Herausforderungen im Bereich der Gestaltung von SDS entstehen. Bisherige Entwicklungsmaßnahmen zur Steigerung der Gebrauchstauglichkeit der Systeme beschränkten sich primär auf das Streben nach einer möglichst fehlerfreien Spracherkennung. In Rahmen der vorliegenden Dissertation wird allerdings das Dialogdesign als Stellschraube für die Nutzerfreundlichkeit diskutiert und die Verringerung von potenziellen Problemen der sprachlichen Mensch-Maschine-Kommunikation im Fahrzeug aus psychologischer Sicht erörtert. Ziel der Arbeit soll sein, aufbauend auf derzeitiger Spracherkennungstechnologie, die Gebrauchstauglichkeit aktueller Systeme durch erwartungskonforme und situationsgerechte Dialoggestaltung zu

erhöhen. Insbesondere der Einstieg in die Sprachbedienung soll erleichtert und Bedienfehler reduziert werden, ohne die Ablenkung von der primären Fahraufgabe zu erhöhen.

Vorrangig wird dabei die Übertragung zwischenmenschlicher Kommunikationsstrategien auf die Mensch-Maschine-Kommunikation untersucht. In Anlehnung an die "Collaborative Theory" (Clark, 1996) werden Optimierungspotentiale aktueller SDS aufgezeigt und Systeme umgesetzt, welche die Prinzipien der menschlichen Dialogführung etablieren. Insbesondere die Integration von umfangreichen Rückmeldeprozessen, die Flexibilisierung des Dialogablaufs und die lexikalische Spiegelung des Nutzerinputs wurden als Ansatzpunkte zur Gestaltung nutzerfreundlicher SDS identifiziert. Im Rahmen von standardisierten Experimenten werden sie auf ihre Transferierbarkeit in den Mensch-Maschine-Dialog überprüft und hinsichtlich der nutzerseitigen Akzeptanz und Gebrauchstauglichkeit evaluiert. Die Erkenntnisse der vorliegenden Arbeit sollen genutzt werden, um Gestaltungsempfehlungen für die Weiterentwicklung automotiver Sprachdialogsysteme abzuleiten.

Die Arbeit untergliedert sich wie folgt: In Kapitel 2 werden die theoretischen Grundlagen aus den Bereichen Mensch-Maschine-Interaktion erläutert sowie eine Analyse des Nutzungskontexts automotiver SDS vorgenommen. Durch beispielhafte Dialoge und der Erläuterung typischer Bedien- und Erkennungsfehler werden Einschränkungen und Potentiale von automotiven Sprachdialogsystemen einander gegenüber gestellt. Aufbauend auf den in Kapitel 3 eingeführten zwischenmenschlichen Kommunikationsstrategien werden in Kapitel 4 anhand einer Statusanalyse Optimierungsbereiche aktueller SDS hergeleitet. Kapitel 5 arbeitet die Fragestellungen der Arbeit heraus. Drei Dialogprinzipien wurden konkret in SDS umgesetzt. Deren empirische Untersuchung wird in Kapitel 6 beschrieben, aus dem Gestaltungsempfehlungen abgeleitet werden. Abschließend ordnet Kapitel 7 die Erkenntnisse dieser Arbeit in die bestehende Forschung ein und gibt einen Ausblick auf künftige Untersuchungen im Bereich automotiver SDS.

2 Grundlagen automotiver Sprachdialogsysteme

In diesem Kapitel werden zunächst grundlegende Begriffe definiert, die für das Verständnis der Arbeit eine bedeutende Rolle spielen. Zu Beginn werden verschiedene Konzepte der Mensch-Maschine-Interaktion (HMI) erläutert. Neben einer Beschreibung multimodaler und adaptiver Schnittstellen wird ein allgemeiner Überblick über die Gebrauchstauglichkeit und die Dialoggestaltung von HMIs gegeben.

Der zweite Teil des Kapitels konzentriert sich auf automotive Sprachdialogsysteme. Nach einer kurzen Definition des Begriffs „Sprachdialogsystem“ wird auf den Aufbau der Systeme eingegangen. Der state-of-the-art aktueller automotiver Sprachdialogsysteme wird dargestellt und die Grundlagen der maschinellen Sprachverarbeitungssysteme erläutert. Dabei werden typische Dialogabläufe und Erkennungs- und Bedienfehler anhand von Beispielen beschrieben. Abschließend werden die Potentiale und Einschränkungen eines Sprachdialogsystems bei der Nutzung im Fahrzeug einander gegenübergestellt.

2.1 Mensch-Maschine-Interaktion

Die vorliegende Arbeit wurde innerhalb des Forschungsgebiets der Mensch-Maschine-Interaktion erfasst. Dieses kann nach Hewitt et. al. (1992) als eine Disziplin definiert werden, die sich mit dem Design, der Implementation und der Evaluation von interaktiven Computersystemen beschäftigt.

Innerhalb dieses Interaktionsraumes kann das Mensch-Maschine-System als ein geschlossener Regelkreis verstanden werden, bei dem Menschen über Schnittstellen Informationen mit Maschinen austauschen. Grundsätzlich kann der Begriff „**Maschine**“ innerhalb dieses Regelkreises unterschiedlichste technische Systeme; von der Mikrowelle bis zum Fahrzeug, bezeichnen. Als **Schnittstellen** zu diesen Maschinen können laut DIN-Norm 9241-110 alle Systemkomponenten bezeichnet werden, die Informationen und Steuerelemente zur Verfügung stellen, welche der Nutzer benötigt, um seine Ziele zu erreichen (Deutsches Institut für Normung, 2006).

Im Fahrzeug herrschen besondere Bedingungen für die Mensch-Maschine-Interaktion, da es sich um einen Nutzungskontext handelt, in dem der Fahrer mit vielfältigen Aufgaben konfrontiert wird. Diese Aufgaben lassen sich nach ihrer Priorität drei Bereichen zuordnen: primäre, sekundäre und tertiäre Fahraufgaben (vgl. Bubb, 2003). Die wichtigste und damit **Primäraufgabe** des Fahrers stellt die Fahrzeugführung im Verkehrsgeschehen dar. Neben der Planung und Navigation, aus der sich die Fahrtroute und die durchschnittliche Geschwindigkeit ableiten lassen, beinhaltet dieses Aufgabenspektrum auch die Ausführung von Fahrmanövern (z. B. Abbiegen oder Überholen) und die Reaktion auf externe Faktoren (z. B. Witterung oder andere Verkehrsteilnehmer). Den

sekundären Fahraufgaben werden jene Reaktionen auf Umgebungsinformationen zugeordnet, die nicht die Längs- oder Querführung des Fahrzeugs adressieren. Als Beispiel können hier die Sichterweiterung durch Anstellen des Fernlichts oder die Kommunikation mit anderen Verkehrsteilnehmern durch Blinken aufgeführt werden. Die verbleibenden **tertiären Aufgaben** stehen nicht in direkter Beziehung zur Fahraufgabe und stellen vorrangig die Komfort-, Unterhaltungs- und Informationsbedürfnisse des Fahrers zufrieden. Zu diesen Tätigkeiten zählen die Bedienung des Radio- oder Navigationssystems und die Interaktion mit angeschlossenen externen Geräten, wie Mobiltelefone (ebd.).

Die vorliegende Arbeit fokussiert auf Interaktionen im tertiären Bereich, deren Priorität den direkten Fahraufgaben unterlegen ist. Deshalb muss die beeinträchtigungsfreie Ausführung der primären Tätigkeiten zu Gunsten der Sicherheit permanent gewährleistet sein und auch bei dem Design von tertiären Schnittstellen im Vordergrund stehen. Nur wenn es sicher möglich ist, sollten sekundäre oder tertiäre Aufgaben durchgeführt werden. Hier können multimodale Schnittstellen ein wichtiger Schritt sein, um Informationen für primäre, sekundäre und tertiäre Aufgaben möglichst gut und sicher zu integrieren (Bengler, 2001). Neben auditiven, visuellen oder haptischen Schnittstellen (Hedicke, 2002) sind gerade im Fahrtkontext auch multimodale Interaktionen denkbar, die im folgenden Kapitel Erläuterung finden.

2.1.1 Multimodale Mensch-Maschine-Interaktionen

Multimodalität bezeichnet die Möglichkeit des Nutzers über verschiedene Arten mit der Maschine zu interagieren bzw. unterschiedliche Modalitäten zum Informationsaustausch zu verwenden. Abbildung 1 verdeutlicht die Interaktionsumgebung des Mensch-Maschine-Systems mit multimodalen Schnittstellen in einem Schaubild.

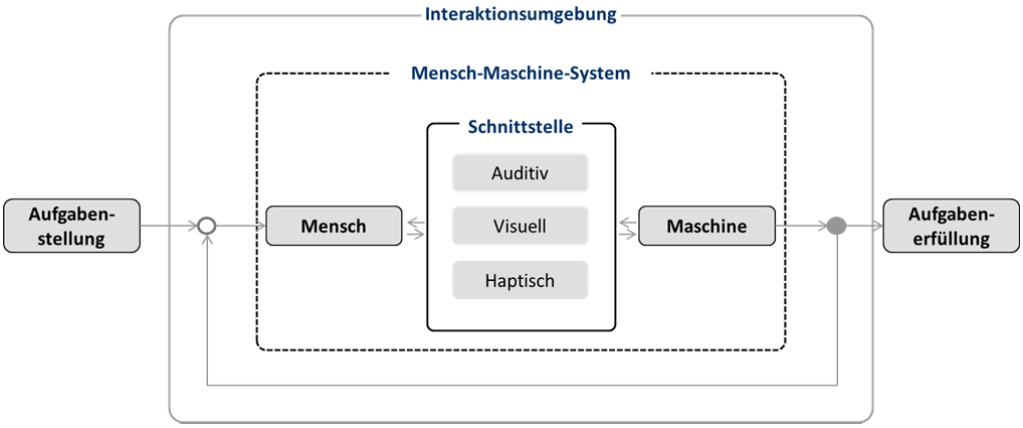

Abbildung 1: Mensch-Maschine-Interaktion über eine Schnittstelle nach Bubb (1993)

Man unterscheidet hierbei zwischen Aktions- und Wahrnehmungsmodalitäten, je nachdem ob eine Modalität zur Ein- oder Ausgabe verwendet wird (Hedicke, 2002). In aktuellen Fahrzeugkonzepten erfolgt die Eingabe des Nutzers häufig über die haptische Modalität (z. B. Drehdrücksteller), während die Ausgabe des Systems häufig über ein visuelles Anzeigesystem erfolgt. SDS bieten die Besonderheit, dass sie die auditive Modalität (Hören/Sprechen) sowohl für die Ein-, als auch die Ausgabe nutzen.

Multimodale Systeme können anhand des vereinfachten Modells von Nigay und Coutaz (1993) nach Hof (2007) klassifiziert werden. Anhand der zeitlichen Abfolge und dem Grad der Informationsfusion werden multimodale Ein- und Ausgaben in das 4-Felder Schema (siehe Abbildung 2) eingeteilt.

Dabei werden auf der zeitlichen Dimension sequentielle und parallele Interaktionen unterschieden. Der parallele Gebrauch erlaubt dem Nutzer die Verwendung mehrerer Modalitäten gleichzeitig, während bei sequentiellen Interaktionen die Nutzung nacheinander geschieht. Die zweite Dimension beschreibt die Verbundenheit der Informationen der verschiedenen Modalitäten. Es wird unterschieden, ob die Informationen komplementär zueinander oder redundant dargeboten werden (Brey & Salmen, 2003).

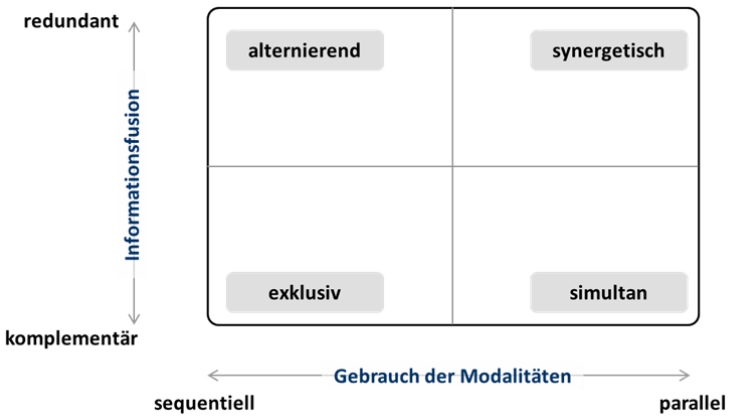

Abbildung 2: Klassifikationsschema multimodaler Systeme (Nigay & Coutaz, 1993)

Somit ergeben sich die folgenden vier Felder für die Klassifikation multimodaler Systeme:

- **Sequentiell-komplementäre Interaktion:** unabhängige Informationen werden in verschiedenen Modalitäten zeitlich aufeinanderfolgend präsentiert
- **Sequentiell-redundante Interaktion:** gleiche Informationen werden aufeinanderfolgend und über verschiedene Modalitäten präsentiert

- **Parallel-komplementäre Interaktion:** verschiedene Informationen werden parallel und über verschiedene Modalitäten präsentiert (z. B. Kartendarstellung des Navigationssystems und Radio)
- **Parallel-redundante Interaktion:** gleiche Informationen werden parallel und über verschiedene Modalitäten präsentiert.

Um sicherzustellen, dass jederzeit die primäre Fahraufgabe mit ihren visuellen und haptischen Anforderungen ausgeführt werden kann, sollten Informationen der tertiären Aufgaben mit parallel-komplementären Charakter in einer anderen Modalität präsentiert werden, um mit der primären Aufgabe möglichst wenig zu interferieren (Wickens, 2002). Damit kommt insbesondere der Sprachbedienung großes Potential zu, auf das in Kapitel 2.3.1 näher eingegangen wird. Ein weiterer Ansatz zur Reduktion der Distraktion tertiärer Aufgaben sind adaptive Schnittstellen, die im folgenden Abschnitt beschrieben werden.

2.1.2 Adaptive Mensch-Maschine-Schnittstellen

Ein wichtiger Aspekt bei der Gestaltung von Mensch-Maschine-Schnittstellen bildet die Adaptierbarkeit. Dabei kann die Anpassung des Systems system-initiiert und nutzer-initiiert geschehen (Dix et al., 2004). Während bei der **nutzer-initiierten Anpassung** die Modifikation der Systemeigenschaften durch den Nutzer selbst durch die Eingabe seiner Präferenzen geschieht, passt sich das System im system-initiierten Fall durch die Analyse von Nutzerinteraktionen automatisch an. In der vorliegenden Arbeit wird nur auf letztere, die **system-initiierte Anpassung** eingegangen und der Begriff der Adaption nur für system-initiierte Anpassungen verwendet.

Welche Kontextmerkmale oder Nutzereigenschaften zur Anpassung herangezogen werden, definiert das sogenannte **Benutzermodell** (BM). Nach Kass und Finin (1988) können diese nach folgenden Charakteristiken eingeteilt werden:

- **Grad der Spezialisierung:** Generische BM betrachten Nutzer als eine homogene Population, wohingegen individuelle BM spezifische Informationen über einen Nutzer enthalten.
- **Modifizierbarkeit:** Findet keine Anpassung des Modells während der Bedienung statt, so spricht man von einem statischen BM. In dynamischen BM fließen dagegen Informationen über den Nutzer noch während der Interaktion ein.
- **Zeitliche Gültigkeit:** Werden die Informationen über den Nutzer im Systemspeicher kumuliert, so bezeichnet man das BM als ein Langzeitmodell. Werden die Informationen allerdings nur während der Bedienung erstellt und anschließend wieder verworfen, so spricht man von einem Kurzzeitmodell.

- **Anwendungsmethode:** Während vorhersagende BM Informationen nutzen, um Verhaltensweisen eines Nutzers zu simulieren und das System entsprechend anzupassen, sammeln beschreibende BM lediglich Informationen über den Nutzer in einer Datenbasis.
- **Anzahl an Agenten:** ein BM kann für eine Person oder Personengruppe ausgelegt sein.

Besonders im Fahrzeug, in dem der Fahrer nicht seine gesamte Aufmerksamkeit zur Ausführung der tertiären Aufgaben aufwenden kann, sollte die Adaption des Systems die Effektivität und Effizienz der Interaktion unterstützen. Gerade die Beschleunigung der Zielerreichung sollte durch die systemseitige Anpassung an die Situation und den Nutzer adressiert werden, um die Ausführung der primären Fahraufgabe möglichst wenig zu beeinträchtigen.

2.1.3 Gebrauchstauglichkeit von Mensch-Maschine-Schnittstellen

Grundsätzlich beschreibt der Begriff „nutzerfreundlich“ die Bedienbarkeit technischer Systeme und bietet ein Maß für die Nutzungsqualität der Mensch-Maschine-Schnittstelle. Synonym werden auch die Begriffe „Bedienbarkeit“ oder „Gebrauchstauglichkeit“ (engl. *Usability*) verwendet.

Das Deutsche Institut für Normung (1998) befasst sich in der Norm 9241 mit „Ergonomischen Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“. Diese enthält Richtlinien für die Interaktion zwischen Mensch und Computer und ist ausdrücklich unabhängig von einer bestimmten Dialogtechnik anwendbar. Speziell die Teile 9241-11 (Anforderungen an die Gebrauchstauglichkeit von Softwareprodukten) und 9241-110 (Grundsätze der Dialoggestaltung) sind für diese Arbeit von Relevanz. Letztere definiert den Begriff der Gebrauchstauglichkeit wie folgt:

„[…] das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele ***effektiv****,* ***effizient*** *und* ***zufriedenstellend*** *zu erreichen“* (ebd., S. 4).

Die Gebrauchstauglichkeit erfordert folglich die Realisierung der drei letztgenannten Leitkriterien, die wie folgt beschrieben werden können:

1. **Effektivität** ist die Genauigkeit und Vollständigkeit der Zielerreichung. Das wichtigste Ziel bei der Nutzung eines interaktiven Systems ist es, die vorliegenden Aufgaben möglichst vollständig und korrekt zu erfüllen. Die Effektivität hängt dabei in hohem Maße von der Funktionalität und Zuverlässigkeit eines Systems ab.

2. **Effizienz** ist das Verhältnis des eingesetzten Aufwands zu der Genauigkeit und Vollständigkeit der Zielerreichung. Hier geht es darum, dass der Nutzer die anstehenden Aufgaben mit der vorhandenen Systemfunktionalität möglichst schnell und mit wenig Aufwand lösen kann. Die

Effizienz lässt sich durch das Verhältnis von prinzipiell nötigen Aktionen (im Idealfall) zu den tatsächlich durchgeführten Aktionen (im Anwendungsfall) beschreiben.

3. **Zufriedenheit** ist die Freiheit von Beeinträchtigungen sowie positive Einstellungen gegenüber der Nutzung des Produkts. Die Zufriedenheit ist eine subjektive Reaktion der Nutzer. Bei besonders positiven Interaktionen spricht man auch von Spaß bei der Nutzung (engl. *Joy of use*).

Beim Entwurf von Softwaresystemen wird die ISO 9241-11 um die Grundsätze der Dialoggestaltung ergänzt (Deutsches Institut für Normung, 2006). Diese sind im Teil 110 der Norm 9241 beschrieben und auf verschiedene Modalitäten anwendbar.

- **Aufgabenangemessenheit:** Ein Dialog sollte zweckmäßig sein und unnötige Interaktionen minimieren. Der Nutzer sollte sein Ziel effizient und effektiv erreichen.
- **Selbstbeschreibungsfähigkeit:** Das System sollte verständlich sein und ausreichend Rückmeldungen oder Erklärungen zur Verfügung stellen.
- **Steuerbarkeit des Dialogs:** Der Dialog sollte nutzerseitig start- und unterbrechbar sein.
- **Erwartungskonformität:** Der Dialog sollte stets konsistent und für den Nutzer nachvollziehbar verlaufen.
- **Fehlerrobustheit:** Bedienfehler des Nutzers werden vermieden oder so abgefangen, dass das Dialogziel erreicht werden kann.
- **Individualisierbarkeit:** Das System kann sich an den Nutzer und die Erfordernisse der Arbeitsaufgabe anpassen (lassen).
- **Lernförderlichkeit:** Der Dialog bedarf nur einer minimalen Anlernzeit. Das System unterstützt den Nutzer beim Erlernen.

Da nur eine beeinträchtigungsfreie Bedienung der Infotainmentsysteme die Ablenkung von der Primäraufgabe in einem akzeptablen Ausmaß erreichen kann, kommt der Gebrauchstauglichkeit bei der Schnittstellengestaltung im Fahrzeug eine zentrale Rolle zu.

Damit sind die zentralen Aspekte der Mensch-Maschine-Interaktion, die auch bei der Gestaltung von SDS relevant sind, eingeführt. Im Folgenden werden nun Sprachdialogsysteme mit ihren besonderen Anforderungen detailliert vorgestellt.

2.2 Definition Sprachdialogsysteme

In der vorliegenden Arbeit soll auf ein Mensch-Maschine-System fokussiert werden, das Sprechen und Hören als primäre Modalitäten zum Informationsaustausch zwischen Nutzer und Ma-

schine verwendet. Nach Möller (1999) wird eine zielorientierte Interaktion zweier Partner als **Dialog** bezeichnet.

Dialog = Sprachliche Äußerungen zwischen mindestens zwei Gesprächspartnern zur Erreichung eines gemeinsamen Interaktionsziels.

Dieser besteht in der Regel aus einer begrenzten Anzahl an sprachlichen Äußerungen, die die Gesprächspartner beitragen. Ein solcher Beitrag von einem der Gesprächsteilnehmer soll im Folgenden als die kleinste Kommunikationseinheit betrachtet und durch den englischsprachigen Begriff „*Turn*“ bezeichnet werden. In einem ausgewogenen Dialog wird jede Äußerung durch einen Sprecherwechsel (engl. *Turn Take*) begrenzt.

Während sich der zwischenmenschliche Smalltalk nur indirekt mit der stark zielorientieren Dialogdefinition nach Möller (ebd.) in Einklang bringen lässt[1], besitzt sie doch vollständige Gültigkeit für die Interaktion mit einem sprachverarbeitenden System. So verfolgt der Nutzer bei dem Austausch mit einem solchen System stets ein kommunikatives Ziel (z. B. Aufbau eines Telefonanrufs, Zieleingabe).

Durch die Fähigkeit des Systems, auf die Nutzereingabe mit einer Systemausgabe zu reagieren, wird ein dialogähnliches Schema zwischen Mensch und Maschine initiiert. Dies ist insbesondere dann der Fall, wenn auf eine Nutzereingabe eine Systemreaktion folgt, die eines weiteren nutzerseitigen Turns bedarf.

Mit Blick auf die **Freiheit der Nutzereingaben** lässt sich die Sprachbedienung grob in zwei Systemarten unterteilen. Neben kommandobasierter Sprachsteuerung existieren natürlichsprachliche Systeme, die auch ganze Sätze und Synonyme bei der Eingabe zulassen. Bei kommandobasierten Systemen sind nur bestimmte Worte zugelassen; Synonyme oder die Eingabe in ganzen Sätzen sind nicht vorgesehen. Der kleine Wortschatz dieser kommandobasierten Systeme führt einerseits zu einer hohen Fehlerrobustheit, andererseits aber auch zu einer großen nutzerseitigen Einschränkung. Die Kommandos müssen häufig erst erlernt werden und bei komplizierten Eingaben ist eine hohe Anzahl von Interaktionsschritten nötig. Der Trend in der Automobilbranche geht daher eher in Richtung natürlich-sprachliche Dialogsysteme. Das sprachverarbeitende System, auf das in dieser Arbeit referenziert wird, soll in Anlehnung an Hamerich (2009) wie folgt definiert werden.

---

[1] Ziel könnte hier sein, Zeit angenehm miteinander zu verbringen.

automotives Sprachdialogsystem = System, welches durch die Verarbeitung von natürlich-sprachlichen Nutzereingaben das Erreichen von nutzerdefinierten, tertiären Zielen im Fahrtkontext ermöglicht.

Um das kommunikative Ziel zu erreichen, muss das SDS Nutzeräußerungen adäquat verarbeiten und sinnvoll darauf reagieren können. Zum detaillierten Verständnis jener Vorgänge werden die Grundlagen der Spracherkennungstechnologie in den folgenden Kapiteln kurz erläutert.

2.2.1 Systemaufbau

Die wichtigsten Komponenten eines Sprachdialogsystem bilden nach Kraiss (2006) und McTear (2002) die Komponenten zur Spracherkennung und zum Sprachverstehen, zum Dialog-Management, zur Sprachgenerierung und -synthese (siehe Abbildung 3).

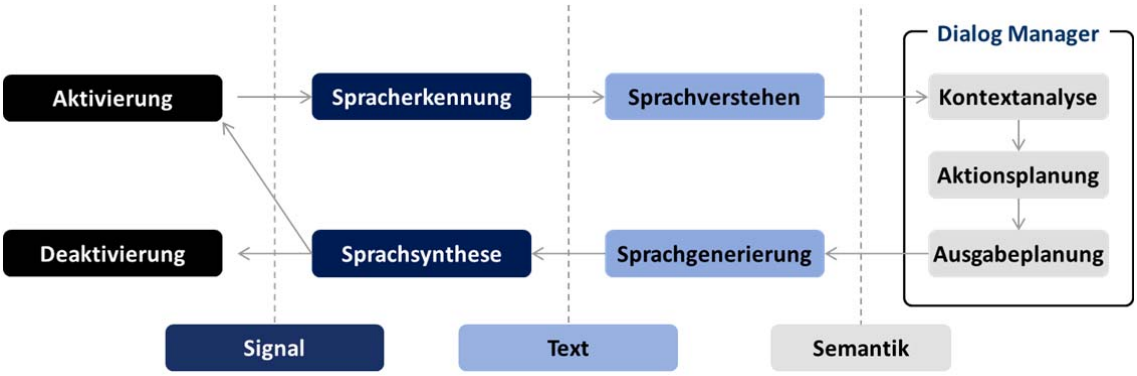

Abbildung 3: Sequenzdiagramm für Sprachdialogsysteme nach McTear (2002)

Im Folgenden sollen die einzelnen Bestandteile eines Sprachdialogsystems kurz erläutert werden.

Um eine permanente Belastung der Rechenleistung des Infotainmentsystems und Fehlerkennungen zu vermeiden, sind SDS im Fahrzeug nicht permanent aktiv. Das Mikrofon muss vom Nutzer gezielt aktiviert werden. Bei den meisten automotiven Systemen wird die Sprachbedienung über einen am Lenkrad verbauten **Push-To-Talk-Knopf** (PTT) aktiviert. Über diesen Knopf lässt sich bei den meisten Systemen der Dialog auch abbrechen. Er stellt damit die einzige haptische Schnittstelle des SDS dar. Die Marke Volkswagen und verschiedene portable Navigationsgeräte mit Sprachfunktion bieten zusätzlich die Möglichkeit an, über einen Softkey auf dem Touchscreen die Sprachaufzeichnung zu starten. Nach dem Tastendruck erfolgt die Sprachauf-

zeichnung über ein oder mehrere Mikrofone. Zur Verbesserung der Aufzeichnung müssen vor der eigentlichen Spracherkennung verschiedene Störgeräusche aus dem Signal gefiltert werden, um dessen Qualität zu verbessern[2]. Insbesondere automotive SDS sind mit hohen **Störgeräuschen**, wie z. B. durch Wind, Scheibenwischer, geöffnete Fenster, Gebläseausströmer oder einer unebenen Fahrbahn, konfrontiert.

Nachdem das Sprachsignal vorverarbeitet wurde, wird es an den **automatischen Spracherkenner** weitergegeben. Das akustische Modell der Spracherkennung, in der Regel gebildet durch Hidden–Markov–Modelle[3] (HMM), wird zur Erkennung von Phonemen verwendet. Alle sprechbaren Kommandos werden zur Zeit der Systemerstellung durch ein G2P (engl. *grapheme to phoneme*) in Phoneme umgewandelt. Die G2P-Technologie erlaubt damit eine automatisierte Phonemisierung von schriftlich vorhandenen Worten im System, die anschließend mit dem bei der Spracheingabe aufgezeichneten Signal, das durch die HMMs in Phonemen vorliegt, verglichen werden kann. Die automatische Spracherkennung bildet somit aus dem Sprachsignal eine diskrete Phonem- bzw. Graphemfolge, indem ein akustisches Modell eingesetzt wird. Ein Sprachmodell oder eine Grammatik generiert dann Worthypothesen. Dabei ist das Ziel, bei einer gegebenen Beobachtung B mit bestimmten Merkmalsvektoren die tatsächlich geäußerte Wortfolge W zu extrahieren. Es wird auf jene Wortfolge zurückgegriffen, welche bei dem gegeben Sprachsignal die größte Wahrscheinlichkeit aufweist. Zur Berechnung dieser Wahrscheinlichkeit findet die Bayes'sche Formel Anwendung:

$$P(W|B) = \frac{P(B|W) \cdot P(W)}{P(\mathrm{B})}$$

Die Wahrscheinlichkeit einer Äußerung gegeben der bestimmten Wortfolge [P(B|W)] stellt in diesem Fall das akustische Modell dar, welches mit Hilfe von Hidden–Markov–Modellen und/oder neuronalen Netzen gebildet wird. Die Wahrscheinlichkeit der Wortfolge [P(W)] bildet das sogenannte Sprachmodell, welches Wissen über die Auftrittswahrscheinlichkeiten von Phonemketten (Reihe von kleinsten bedeutsamen Lauteinheiten) und Hypothesenketten (Wortkombinationen im Satzbau) enthält. Aus der Multiplikation der Übereinstimmungswahrscheinlichkeiten mit ihren Auftrittswahrscheinlichkeiten ergibt sich der Übereinstimmungsgrad eines jeden Datenbankein-

---

[2] Um detaillierten Einblick in die Verfahren der Echo- und Störreduktion zu erhalten, sei auf Hänsler und Schmidt (2004) und Nordholm, Claesson & Grbic (2001) verwiesen.

[3] Die Grundlagen der Sprachsignalerkennung und eine Übersicht zu HMMs in der Spracherkennung sind bei Jelinek (1998) und Schenk & Rigoll (2010) ausführlich dargestellt.

trags bezüglich der erfassten Spracheingabe. Diese Wahrscheinlichkeit wird auch als **Konfidenzmaß** beschrieben und quantifiziert die Relevanz des potentiell zur Spracheingabe passenden Datenbankeintrags.

Zur Modellierung sprachlicher Einheiten oberhalb der Wortebene dienen Syntaxmodelle. Das Sprachmodell errechnet die wahrscheinlichste syntaktische Kette und reichert diese ggf. mit Metainformationen an (IDs für Namen, Kommandos, etc.). Danach wählt ein semantischer **Parser** (Bedeutungsmodell) die im aktuellen Dialogschritt relevanten Bausteine der Spracheingabe aus. Die Syntax einzelner Phrasen oder ganzer Sätze kann entweder durch einen grammatikbasierten Ansatz (engl. *grammar based model*; GBM) oder ein statistisches Sprachmodell (engl. *statistical language model;* SLM) abgebildet werden. Bei Letzterem werden die Wahrscheinlichkeiten der Wortfolgen durch stochastische Prozesse identifiziert. Das SLM wird zuvor auf der Basis großer Textkorpora maschinell trainiert. Im Rahmen von grammatikbasierten Ansätzen werden die möglichen Wortfolgen in einem vereinbarten Grammatikformat manuell spezifiziert. Nach der regelbasierten Verarbeitung durch die Grammatik setzt der semantische Parser an. Beide Technologien werden abhängig vom Anwendungsbereich eingesetzt: Bevorzugen Nutzer eine stark kommandobasierte Eingabe, so ist das GBM dem SLM überlegen. Dagegen zeigen SLMs bei einer freieren und natürlich-sprachlicheren Eingabe durch den Nutzer eine größere Leistungsfähigkeit (Gorrell, 2003; Hockey, et al., 2003).

Die Ergebnisse des Spracherkennungsprozesses werden durch den **Dialog Manager** (DM) interpretiert. Er stellt das Herz und die Schnittstelle des Sprachdialogsystems zu den Sprachein- und -ausgabegeräten und weiteren Systemkomponenten (z. B. Datenbank) dar. Unter Berücksichtigung des Kontextes und der Dialoghistorie, werden die Spracheingaben des Nutzers zunächst analysiert und bewertet. Abhängig vom Analyseergebnis entscheidet der DM über weitere Dialogschritte und leitet Aktionen innerhalb eines Dialogzustands ein. Ist das Ergebnis der Spracherkennung beispielweise nicht eindeutig interpretierbar, so wird typischerweise eine Trefferliste oder eine systemseitige Rückfrage ausgelöst. Ist das Ergebnis eindeutig, so wird der Datenbankeintrag ausgegeben oder eine entsprechende Aktion ausgeführt. Die Zustandsmodellierung erfolgt dabei mittels einer Beschreibungssprache wie VoiceXML oder der Generic Dialogue Modeling Language (GDML) (Hanrieder & Hamerich, 2004).

Die sprachlichen Rückmeldungen des Systems (engl. ***Prompts***) erfolgen über die Bordlautsprecher des Fahrzeugs. Die Ausgabe erfolgt dabei entweder per bereits vorab aufgenommener Sprachaufzeichnungen menschlicher Sprecher (engl. *pre-recorded speech*) oder per synthetischer Sprachausgaben (engl. *text to speech*, TTS).

Vor dem Hintergrund dieser technologischen Rahmenbedingungen sollen die Dialoge zwischen Mensch und Maschine im Fokus der vorliegenden Arbeit stehen. Um die Charakteristiken dieser

Interaktionen zu verdeutlichen, werden im Folgenden die Verbreitung und der Funktionsumfang derzeitiger automotiver Sprachdialogsysteme dargestellt. Anhand typischer Dialogabläufe sollen Herausforderungen der Schnittstellengestaltung verdeutlicht werden.

2.2.2 Funktionsumfang automotiver SDS

Betrachtet man den heutigen Markt, so stößt man vielerorts auf Anwendungen, die bereits von Sprachbedienung Gebrauch machen. Etabliert haben sich diese Systeme bisher vorrangig im Bereich der Telefonauskunftssysteme, in Mobilfunkgeräten und im Fahrzeug (vgl. Hamerich, 2009). Neben dem Originalzubehör der Automobilhersteller halten SDS dabei in zunehmendem Maße auch Einzug in Nachrüstgeräte, wie z. B. portable Navigationsgeräte (Hanrieder, 2004).

Die herkömmlichen Systeme, wie sie derzeit mit dem MMI 3G+ bei Audi, der RNS510 bei Volkswagen, dem iDrive bei BMW und dem COMANDSystem mit Linguatronic bei Mercedes in Serie sind, stützen sich hauptsächlich auf die fest verbaute Elektronik im Auto. Mit Hilfe der Sprachbedienung lassen sich verschiedene Kontexte wie Navigation, Medien und Adressbuch bedienen. Neben kontextsensitiven Kommandos (z. B. *hinfahren*) funktionieren Basisbefehle wie *Hilfe* oder *Abbrechen*. Neben diesen herkömmlichen Funktionen unterstützt das derzeitige Ford SYNC System ein Vorlesen und Diktieren von Textnachrichten.

Die meisten der aktuellen, automotiven SDS ermöglichen dabei ausschließlich eine Eingabe bestimmter Kommandos und reagieren mit konstanten Sprachausgaben. Dabei sind sie jedoch **sprecherunabhängig**[4] gestaltet und können ohne eine explizite Anlernphase des Erkenners verwendet werden. Zu dem heutigen Zeitpunkt stellt jedes dieser sprachverarbeitenden Systeme im Fahrzeug eine ergänzende Bedienmodalität dar. Sämtliche Funktionen können auch haptisch ausgeführt werden.

Die **Dialoge** mit Sprachdialogsystemen weisen eine große Bandbreite bezüglich ihrer Komplexität auf. Das Kontinuum der Nutzereingabe kann dabei von kurzen Ein-Wort-Bestätigungen bis hin zu umfangreichen multi-slot-Eingaben wie bei der Navigationszieleingabe reichen. Anhand einiger exemplarischer Funktionalitäten sollen nun einige typische Dialogabläufe dargestellt werden.

[4] Während sprecherabhängige Spracherkenner vom Nutzer vor der Verwendung auf die Besonderheiten der eigenen Aussprache trainiert werden müssen, können sprecherunabhängige Spracherkenner ohne eine vorhergehende Trainingsphase sofort verwendet werden und bieten dabei auch die Gelegenheit, häufig wechselnde Nutzer zu verstehen.

Die Nutzereingabe und Systemausgabe bilden dabei ein dialogähnliches Schema (vgl. Kap. 2.2). Eine bekannte Funktionalität ist die **Wahl einer Telefonnummer** per Sprache. Ein Beispieldialog für diesen Fall ist im Folgenden dargestellt:

Nutzer	<Aktivierung über den Push-To-Talk-Knopf>
System	<Beep>
Nutzer	Nummer wählen
System	Bitte sprechen Sie die Nummer.
Nutzer	0-5-3-6-1
System	0-5-3-6-1 und weiter?
Nutzer	9-1-6-5-3
System	9-1-6-5-3 und weiter
Nutzer	Wählen.
System	Die Nummer 0-5-3-6-1-9-1-6-5-3 wird gewählt.

Der Dialog mag mit den Wiederholungen der Nutzereingabe lang erscheinen, dennoch wird diese Strategie der Rückversicherung häufig verwendet, um dem Nutzer frühzeitig die Möglichkeit zur Korrektur zu bieten. Diese wird in Kap. 2.2.4 noch näher erläutert. Durch eine Bluetooth-Schnittstelle bieten aktuelle Infotainmentsysteme dem Nutzer die Möglichkeit, ein Mobiltelefon mit dem Fahrzeug zu verbinden. Dessen Kontakte werden bei erfolgreicher Kopplung mit dem Fahrzeug synchronisiert und damit sprachbedienbar. Damit wird die Verwendung des Telefons stark vereinfacht und der Zugriff auf Kontakte verkürzt. Ein solcher **Namenwahldialog** könnte wie folgt aussehen:

Nutzer	Anrufen bei Helga Otto.
System	Den Nummerntyp bitte.
Nutzer	Mobil.
System	Möchten Sie Helga Otto mobil anrufen?
Nutzer	Ja.
System	Es wird gewählt.

Ein weiteres, großes Anwendungsfeld der Sprachbedienung im Fahrzeug ist die **Navigationszieleingabe**. Da es allein in Deutschland über 68.000 verschiedene Orte (darunter 32-mal den Ort Neustadt) gibt, stellt die Ganzworteingabe aufgrund der Vielzahl der möglichen Eingaben und deren phonetische Ähnlichkeit eine große Herausforderung dar.

So war es erst 2003, mit dem Erscheinen der Mercedes-Benz E-Klasse, erstmalig möglich die 600 größten Orte eines Landes zur Navigation als Wort einzusprechen, anstatt sie buchstabieren zu müssen. 2009 ermöglichte BMW dem Nutzer eine sogenannte "*one-shot*"-Spracheingabe. Dies bezeichnet eine ganzheitliche Spracheingabe des Navigationsziels, bestehend aus Stadt, Straße und Hausnummer, innerhalb einer Äußerung. Damit können ganze Adressen in einem Satz eingegeben werden und müssen nicht länger kategorienweise eingesprochen werden. Seit der Einführung des A8 ist dies auch im Audi MMI 3G+ System möglich. Ein exemplarischer Dialog ist im Folgenden dargestellt:

Nutzer	Zieleingabe
System	Zieleingabe. Bitte nennen Sie die Stadt, die Straße und die Hausnummer.
Nutzer	Braunschweig, Rebenring, 12.
System	(Anzeige der Ergebnisliste). Bitte nennen Sie die Zeile oder sagen Sie Navigation starten.
Nutzer	Navigation starten.
System	Route wird berechnet.

Um das Problem der ähnlichen (oder gleichen) Ortsnamen und Verwechslungen in den Griff zu bekommen, präsentiert das System dem Nutzer eine Liste der wahrscheinlichsten Ergebnisse (engl. *n-best list*) zur Auswahl. Häufig kann die Auswahl aus dieser Disambiguierungsliste per Sprache oder auf haptischem Weg geschehen. Wird der Telefonanruf oder die Zielführung gestartet, endet der Sprachdialog und die Aktion wird ausgeführt.

Aktuelle SDS haben nur einen **begrenzten Zeitraum**, in dem eine Nutzereingabe zulässig ist. So ist es bei den meisten Systemen nicht möglich, eine Eingabe zu tätigen, während das System selbst spricht. Das sogenannte *Barge-In* (engl. für ins Wort fallen), welches dem Nutzer ermög-

licht das System zu unterbrechen, erfordert, dass das SDS während einer Sprachausgabe in der Lage ist eine nutzerseitige Eingabe zu verarbeiten. In den meisten automotiven Systemen ist lediglich der sog. *Pseudo-Barge-In* umgesetzt: Der Nutzer muss, um das System zu unterbrechen, erneut die Sprechtaste drücken.

Auch in absehbarer Zukunft ist nicht mit der perfekten maschinellen Spracherkennung zu rechnen. Erklären lässt sich dies durch die hochkomplexen Vorgänge des Spracherkennens und -verstehens. Auch für eine einfache Konversation werden umfangreiches Wissen und Fähigkeiten benötigt (Phonetik, Phonologie, Morphologie, Syntax, Semantik, Pragmatik). Vergegenwärtigt man sich diese Komplexität und Vielfältigkeit, wird deutlich, dass auch in aktuellen Systemen noch kein natürlich-sprachlicher Zugang zu Daten des Infotainmentsystems realisiert werden kann. Das folgende Kapitel zeigt Probleme der heutigen maschinellen Spracherkennung auf und führt die häufigsten Erkennungs- und Bedienfehler ein.

2.2.3 Typische Erkennungs- und Bedienfehler

Zunächst sollen zwei Problembereiche aufgeführt werden, die auch in der zwischenmenschlichen Kommunikation zu Verständnisschwierigkeiten führen können und Erkennungsprobleme begünstigen:

- **Segmentierungsproblem:** Da gesprochene Sprache kontinuierlich ist, sind sprachliche Einheiten häufig auch über Wortgrenzen hinweg nicht scharf voneinander abgegrenzt. Wortgrenzen sind damit nicht eindeutig zu identifizieren und werden häufig nur durch Vorwissen erkannt. Menschen erleben dies vor allem, wenn sie eine Fremdsprache hören, derer sie nicht mächtig sind. Ähnlich wie ein Hörer muss ein Spracherkenner Einheiten im Sprachsignal entdecken, die Zugriff auf das mentale Lexikon erlauben.
- **Variabilitätsproblem:** Eine Herausforderung bietet die natürliche Variabilität gesprochener Sprache. Selbst wenn dieselbe Person eine Sprachäußerung zweimal in Folge produziert, sind diese nicht gleich (Harris, 2005). Diese Variabilität tritt jedoch auch zwischen Personen auf. So unterscheiden sich Sprachsignale unter anderem nach Dialekt, Geschlecht, Gesundheitszustand oder emotionalen Zustand.

Beide Probleme erhöhen die Wahrscheinlichkeit von Fehlererkennungen durch das System.

Zur Veranschaulichung möglicher Probleme in der Interaktion mit einem SDS dient Abbildung 4. In Anlehnung an Morgan et al. (2001) wurden die häufigsten Fehler drei Hauptkategorien zugewiesen. Ferner wurde verdeutlicht, ob sie auf ein Fehlverhalten des Systems oder des Nutzers zurückgehen.

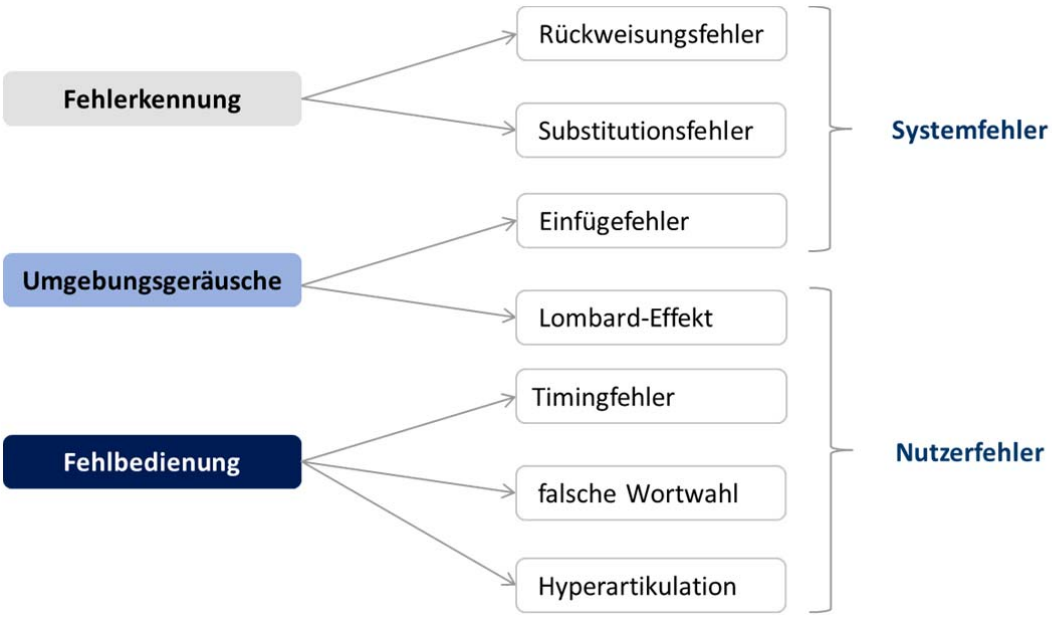

Abbildung 4: Fehlerkategorien eines SDS

Fehlerkennungen (engl. *Recognition Errors*; FE) können in vielerlei Arten auftreten. Sogenannte **Rückweisungsfehler** treten auf, wenn ein System ein Kommando auch dann falsch versteht, wenn dieses in seinem Wortschatz enthalten ist. Diese Fehler führen häufig dazu, dass eine hohe nutzerseitige Unsicherheit über die Systemkompetenzen besteht. Nutzer sind sich nicht sicher, ob sie die falschen Kommandos verwenden oder nur schlecht verstanden werden.

Darüber hinaus können auch **Substitutions- bzw. Ersetzungsfehler** beobachtet werden. Ein eingegebenes Kommando wird fälschlicher als ein anderes Kommando verstanden.

Ein Beispieldialog dafür ist im Folgenden skizziert:

Nutzer	Ich möchte Julia Hahn zu Hause anrufen.
System	Möchten Sie Linda Huber zu Hause anrufen?
Nutzer	Nein, Julia Hahn.

In der automotiven Umgebung kommen darüber hinaus noch störende Umweltbedingungen hinzu, die die Erkennung erschweren. **Umgebungsgeräusche** (engl. *Environmental Errors*), wie Lärm, Echos oder Gespräche unter anderen Fahrgästen können zu sogenannten **Einfügefehlern** führen. Das System interpretiert dann diese Störgeräusche als ein Teil des Sprachsignals und erkennt daraus fälschlicherweise Kommandos (Schmandt, 1994, S. 160). Dabei können wieder-

holt auftretende Umweltgeräusche derzeit signaltechnisch und softwaretechnisch gut herausgefiltert werden. Das Trennen menschlicher Stimmen bleibt dagegen problematisch.

Umgebungsgeräusche führen auch zu dem sogenannten **Lombard-Effekt** (Junqua, 1993). Bei erhöhten Umgebungslautstärken erhöhen Sprecher nicht nur die Lautstärke ihrer Spracheingabe, sondern unwillkürlich auch ihre Tonlage. Da hohe Frequenzen sich über weite Distanzen besser transportieren lassen und sich deutlicher von dem Umgebungsgeräuschpegel abheben, trägt dies zu der Effektivität zwischenmenschlicher Gespräche bei hohen Geräuschpegeln bei. Allerdings stellt der Lombard-Effekt eine häufige Ursache für Fehlerkennungen von SDS dar, da akustische Modelle tendenziell eher mit Sprachdaten erstellt werden, die unter Normalbedingungen aufgenommen wurden. Damit stellt die situationsgerechte Anpassung des Nutzers bereits ein nicht intendiertes Fehlverhalten dar und kann somit als Eingabe- oder Bedienfehler gelten. Laut Mann (2010) stellen die **Fehlbedienungen** die häufigste Problemquelle bei der Interaktion mit SDS dar.

Zu dieser Fehlerkategorie gehören unter anderem eine zu frühe Spracheingabe, falsche Wortwahl oder eine zu späte bzw. keine Eingabe. Letzteres (engl. *no input)* liegt vor, wenn der Nutzer mit seiner Spracheingabe zu lange wartet. Da der Spracherkenner nach einem vorkonfiguriertem Zeitintervall wieder geschlossen wird, tritt in diesem Fall ein sogenannter no input-Fehler auf. Dieser gehört damit, genau wie die zu frühe Spracheingabe, zu der Unterkategorie der nutzerseitigen **Timingfehler**. Der Bedienfehler des zu frühen Sprechens (engl. *talking too early,* TTE-Fehler) beschreibt das Verhalten des Nutzers, eine Spracheingabe zu tätigen, bevor der Audiokanal für den Spracherkenner geöffnet wird. Häufig tritt er durch ein Übersprechen bzw. Unterbrechen der Systemausgabe auf. In diesen Fällen kann meist der Anfang der Äußerung nicht richtig verarbeitet werden, was eine Fehlerkennung wahrscheinlich macht. Häufig bemerkt der Nutzer aufgrund mangelnden Feedbacks dieses Erkennungsproblem nicht und kann ihm keine Ursache zuweisen. Dann sind weitere Turns nötig, die den Dialogverlauf korrigieren und unnötig verlängern (Hamerich, 2009).

Die **falsche Wortwahl** wird auch häufig mit der Abkürzung OOV-Fehler (engl. *out of vocabulary*) bezeichnet. In diesem Fall hat der Nutzer ein Kommando eingesprochen, das nicht im Wortschatz des Sprachdialogsystems vorhanden ist und deshalb nicht erkannt wird. Das Verwenden eines out of vocabulary-Wortes kann allerdings auch zu einem Substitutions- bzw. Ersetzungsfehler führen, da der Spracherkenner versucht die Eingabe auf das vorhandene Vokabular abzubilden und das phonetisch ähnlichste Wort vorschlägt. Der OOV-Fehler stellt damit eine Kombination aus Fehlbedienung durch den Nutzer und Systemfehler dar. Eine stetige Vergrößerung des

Vokabulars kann nicht als alleinige Lösung jener Problematik bestehen. Denn ein Anstieg der Wortschätzgröße benötigt nicht nur zusätzlichen Speicherplatz, sondern bedeutet auch ein Anstieg des Verarbeitungsaufwandes und des Fehlerkennungsrisikos[5].

Erschwerend kommt hinzu, dass Nutzer in der Interaktion mit einer Maschine versuchen, besonders deutlich zu sprechen und insbesondere im Problemfall zu einer Überbetonung ihrer Spracheingaben tendieren. Wenn Sprecher glauben, dass der Empfänger sie nicht versteht, passen sie ihr Sprachverhalten in mehreren Aspekten an. Sie sprechen lauter, langsamer und überdeutlich (Stent et al., 2008). Diese automatisierte Strategie der Überbetonung bzw. **Hyperartikulation** ist zwischenmenschlich äußerst hilfreich (Cutler & Butterfield, 1990; Picheny et al.,1985) und kann die Verständlichkeit von Sprache, insbesondere bei Gesprächen mit Muttersprachlern erhöhen (Bradlow & Bent, 2002). In der Interaktion mit einem automatischen Spracherkennungssystem ist sie allerdings wenig hilfreich und kann zu höheren Fehlerraten führen (Shriberg et al., 1991; Soltau & Waibel, 1998; Wade et al., 1992). Da state-of-the-art Spracherkenner mit flüssig eingesprochener Sprache trainiert sind, steigt die Wahrscheinlichkeit einer Fehlererkennung bei hyperartikulierten Eingaben. Huber et al. (1998) konnten zeigen, dass die Worterkennungsrate von Spracherkennern, die mit normaler Sprache trainiert sind, durch emotionale Spracheingaben drastisch reduziert wird. Auch Levow (1998, S. 737) belegte, dass die Fehlerwahrscheinlichkeit der Spracherkenner von 16% auf 44% bei wiederholten Eingaben steigt, da Nutzer dazu tendieren die wiederholten Eingaben hyperartikuliert wieder zu geben. Damit stellt die Hyperartikulation eine nutzer-initiierte Quelle für Systemfehler dar.

Es lässt sich schließen, dass mannigfaltige Faktoren die Reliabilität der Spracherkennung reduzieren können. Störgeräusche, Überlappungen verschiedener Sprecher, umfangreiche Systemwortschätze, phonologische Ähnlichkeit des Vokabulars oder komplexe Nutzeräußerungen begünstigen Fehlerkennungen. Werden Nutzern darüber hinaus die Systemgrenzen nicht deutlich vermittelt, treten die oben beschriebenen Fehlbedienungen auf. Die Ursache der Timing- oder OOV-Fehler sollten jedoch nicht ausschließlich im Nutzer verortet werden, sondern auch als Mangel der Dialoggestaltung Beachtung finden.

Eine der wichtigsten Stellgrößen, um die Häufigkeit von Erkennungs- bzw. Bedienfehlern zu reduzieren, stellt somit das Dialogdesign dar. Es prägt maßgeblich den (beobachtbaren) Dialogablauf und soll im folgenden Kapitel beschrieben werden.

---

[5] Furnas (1987; Brennan, 1998) führte dies als das "Vocabulary problem" ein.

2.2.4 Dialogdesign

Im Rahmen des Dialogdesigns werden sowohl die groben Linien eines Dialogs, wie die Dialogspezifikation, als auch die Details, wie die Wortwahl für Systemäußerungen, das Timing und die Definition der Grammatik bearbeitet (Hamerich, 2009).

Im Rahmen der grundlegenden Spezifikation lassen sich maßgeblich zwei Arten der Dialogführung definieren, die sich vor allem darin unterscheiden, bei wem die Dialoginitiative liegt.

1. **System-initiiert**: Nach der Aktivierung bestimmt das Dialogsystem durch schrittweises und gezieltes Nachfragen den Dialog.

2. **Nutzer-initiiert:** Diese Systeme überlassen dem Nutzer durch geringe Vorgaben bei der Eingabe die Kontrolle.

Frühere SDS wiesen zumeist eine starre Struktur auf und ließen pro Dialogschritt nur eine begrenzte Anzahl an Eingaben zu. Die Dialoge waren zumeist system-initiiert und werden innerhalb eines Werbespots karikiert:

"Interessieren Sie sich für unsere Bananen, sagen Sie Bananen, interessieren Sie sich für unsere Äpfel sagen Sie Äpfel." (Yello Strom, 2008)

Da der Nutzer häufig nur mit einzelnen Informationen oder Ja/Nein auf geschlossene Fragen antworten muss, kann durch dieses Vorgehen allerdings die Variabilität der Nutzereingaben verringert und damit die Erkennungsgenauigkeit erhöht werden. Insbesondere bei unerfahrenen Nutzern eignet sich diese Art von geführten Dialogen um Fehlerkennungen und Unsicherheiten zu reduzieren. Dagegen weist die nutzer-initiierte Dialoggestaltung den Nachteil auf, dass die Möglichkeit zur freien Formulierung in kaum vorhersehbaren Äußerungen resultieren, die zu einer schlechteren Erkennungsleistung führen können (vgl. Cohen et al., 2004; McTear, 2004). Diese Variante eignet sich also besonders für erfahrene Nutzer, die bereits eine Vorstellung der Systemkompetenz aufgebaut haben.

Aufgrund dieser Vor- und Nachteile beider Ansätze, verwenden aktuelle SDS häufig eine Kombination der beiden Strategien, die **gemischt-initiierte Dialogführung**. So bieten sie z. B. einen nutzerzentrierten Einstieg ("Wie kann ich Ihnen helfen?"), zeigen aber bei der konkreten Aufgabenlösung eine stärkere Systeminitiative ("Den Nummerntyp bitte."). Dies bietet den Vorteil, dass die Strategie je nach Nutzereingabe angepasst werden kann und damit flexibel auf die Bedürfnisse der Nutzer eingeht. Die beiden folgenden Dialogbeispiele zeigen eine Verortung der Initiative in Abhängigkeit des Umfangs der ersten Nutzereingabe.

	Nutzer-Initiative	**System-Initiative**
Nutzer	Ich möchte Helga Otto mobil anrufen.	Anrufen.
System	Möchten Sie Helga Otto mobil anrufen?	Wen möchten Sie anrufen?
Nutzer	Ja.	Helga Otto.
System	Es wird gewählt.	Den Nummertyp, bitte.
Nutzer		Mobil.
System		Helga Otto mobil. Es wird gewählt.

Die Beispieldialoge zeigen, dass die Reihenfolge der Dialogschritte nicht festgelegt ist und je nach Vollständigkeit der Nutzereingabe variieren kann. Das System übernimmt in diesem Fall erst die Initiative, wenn notwendige Informationen zur Ausführung der Aktion fehlen. Der Nutzer kann allerdings durch eine multi-slot Eingabe, die alle relevanten Informationen enthält, den gesamten Dialogablauf stark verkürzen. Diese Art des gemischt-initiierten Dialogdesigns bietet eine hohe Flexibilität im Dialogablauf und begünstigt die Zielerreichung.

Weiterentwicklungen wie z. B. die Arbeiten von Litman und Pan (2002) beschäftigen sich damit, die Initiative **adaptiv** an die Erkennungsgüte anzupassen. Da für sie die nutzer-initiierte Dialogführung der menschlichen Kommunikation am ähnlichsten ist, stellt sie die Ausgangsstrategie dar. Wird im Lauf der Interaktion eine hohe Fehlererkennungsrate registriert, so übernimmt das System die Initiative und erlangt damit eine höhere Robustheit.

Ein weiterer Aspekt des Dialogdesigns, der für diese Arbeit von Bedeutung ist, liegt in dem **Umgang des Systems mit Ergebnissen** des Spracherkenners. Wie bereits in Kapitel 2.2.1 erläutert, besteht die Spracherkennung im Wesentlichen aus einem wahrscheinlichkeitsbasierten Abgleich zwischen dem erkannten Sprachsignal mit einer Datenbank. So besteht das Erkennungsergebnis aus einer Liste von erkannten Phrasen, die jeweils mit einem Konfidenzwert (der Übereinstimmungswahrscheinlichkeit mit einem Datenbankinhalt) versehen sind. Durch das Festlegen einer Konfidenzschwelle entscheidet der Entwickler zunächst welche Erkennerergebnisse dem Nutzer präsentiert werden. Die systemseitige Interpretation dieser Ergebnisse bietet Gestaltungsspielraum. Nach Litman und Pan (1999, 2002) können dabei drei Strategien Anwendung finden:

1. **Defensive Interpretation:** Der Nutzer muss jede erkannte Eingabe explizit bestätigen. Diese Art der expliziten Verifikation erhöht die Robustheit, verlängert allerdings den Dialog.

2. **Offensive Interpretation**: Ein erkannter Begriff wird so lange als korrekt angenommen, bis der Nutzer widerspricht. Durch eine erneute Präsentation des Erkennungsergebnisses wird dem Nutzer einerseits die Gelegenheit gegeben eine Korrektur einzuleiten oder im Dialog fortzufahren. Dies stellt eine effizientere Strategie der Verifikation dar.

3. **Aggressive Interpretation**: Das System fragt nicht nach einer Bestätigung und führt Aktionen sofort aus.

Die letztgenannte Strategie gilt als die menschenähnlichste, findet aufgrund einer zu hohen Gefahr von falschen Systemreaktionen jedoch nur selten Anwendung. Im Grunde stellt sie eine Form der offensiven Interpretation ohne Korrekturmöglichkeit dar. Auch hier sind Kombinationen der Strategien denkbar. Wie im anschließenden Beispieldialog dargestellt, kann bis zu einer finalen Bestätigungsanfrage eine aggressive Interpretation verfolgt werden:

Nutzer	Helga Otto
System	Was möchten Sie tun?
Nutzer	Anrufen
System	Den Nummerntyp bitte.
Nutzer	Mobil
System	Möchten Sie Helga Otto mobil anrufen?

Nachdem der Funktionsumfang, die Probleme und Gestaltungsalternativen aktueller SDS aufgezeigt wurden, sollen im folgenden Kapitel die Potentiale und Einschränkungen von Sprachbedienung im Fahrtkontext diskutiert werden.

2.3 Potentiale und Einschränkungen automotiver SDS

Sprache als Bedienmodalität bietet im Fahrtkontext entscheidende Vorteile. Einerseits konnte umfangreiche Forschung belegen, dass sich die Interaktion mit SDS im Vergleich zu visuell-haptischen Bedienschnittstellen nicht negativ auf die Primäraufgabe der Fahrzeugführung auswirkt (im Überblick, s. Vollrath & Totzke, 2000). Andererseits kann der Zugang zu dem Infotainmentsystem des Fahrzeugs über die Sprache als eine natürliche und intuitive Kommunikationsform bewertet werden. Darüber hinaus bieten SDS eine vielversprechende Möglichkeit das stets wachsende Funktions- und Datenangeboten bedienbar zu machen. Dennoch kann eine deutliche Diskrepanz zwischen Kundenerwartungen und technischer Realität festgestellt werden. Obwohl in den letzten Jahrzehnten deutliche Verbesserungen der Spracherkennungstechnologie erreicht

werden konnten, sind aktuelle SDS immer noch weit von natürlich-sprachlicher Kommunikation, wie sie mit zwischenmenschlichen Gesprächspartnern möglich ist, entfernt und können die Erwartungen der Nutzer häufig nicht erfüllen.

Im Folgenden sollen zunächst die Vor- und Nachteile der Sprachbedienung ausführlicher erläutert werden. Dabei soll insbesondere die Bedeutung von Sprachbedienung im automotiven Kontext und der Vergleich zu anderen Bedienmodalitäten beleuchtet werden. Darauffolgend soll die Notwendigkeit von Sprachbedienung bei wachsendem Funktionsumfang und die nutzerseitige Akzeptanz der Schnittstellen anhand von empirischen Erhebungen diskutiert werden.

2.3.1 Sprache als bedingt sichere Bedienmodalität

Studien konnten belegen, dass die Verarbeitung akustischer Informationen im Fahrzeug nicht zwingend zu einer Beeinträchtigung der Primäraufgabe führen. So hat Merat (2003) nachweisen können, dass eine auditive Zweitaufgabe während der Fahrt das Fahrverhalten nicht systematisch beeinträchtigt. Auch bei Röder und Wank (2011) konnte gezeigt werden, dass eine phonologische Nachsprechaufgabe verglichen zur Referenzfahrt keine signifikante Verschlechterung der Fahrleistung zeigte. Die geringere Interferenz der Sprachbedienung mit der Fahraufgabe lässt sich durch das **Multiple Ressourcen Modell** von Wickens (2002) erklären (siehe Abbildung 5). Nach diesem Modell hängt das Ausmaß der Interferenz zweier Aufgaben davon ab, ob beide Aufgaben die gleichen Ressourcen beanspruchen (Wickens, 2002). Die Vorstellung von multiplen und spezialisierten Verarbeitungssystemen führt zu der zentralen Annahme, dass die kritische Determinante für die Zweitaufgabenperformanz die Aufgabenähnlichkeit ist. Aufgaben, die denselben Bereich der Matrix ansprechen, interferieren mehr als Aufgaben, die unterschiedliche Felder besetzen.

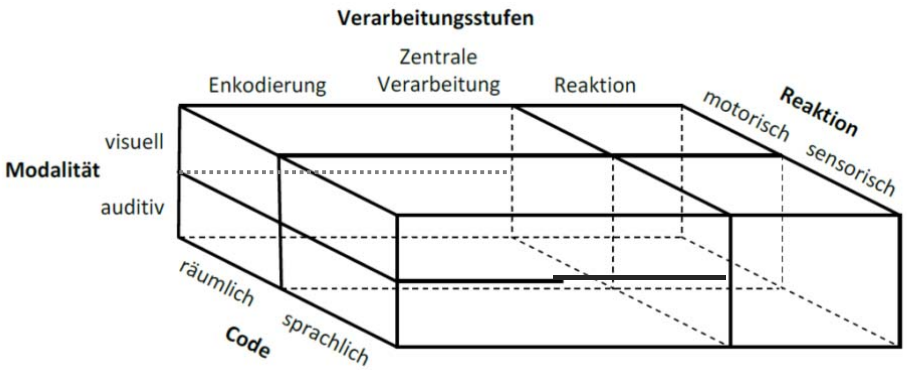

Abbildung 5: Multiples Ressourcen Modell nach Wickens (2002)

Da die Informationsaufnahme während der Fahrt primär über die visuelle Modalität erfolgt und die Reaktionen vorrangig motorischer Natur sind, sind die visuellen Felder der Inputmodalität und die manuellen der Reaktionsmodalitäten belegt. Daher führen insbesondere visuelle Nebenaufgaben zu einer hohen Interferenz und werden als eine der häufigsten Unfallursachen diskutiert (Wierwille & Tijerina, 1995). Dagegen bieten sprachliche Nutzerschnittstellen den Vorteil, eine hand- und augenfreie Interaktion zu ermöglichen und damit die Ressourcen, die durch die Fahraufgabe bereits gebunden sind, nicht zusätzlich zu belasten. Gerade die auditive Modalität besitzt während der Fahrt noch freie Kapazitäten und legt somit eine Erweiterung der visuell-haptischen Schnittstellen durch die sprachliche Modalität nahe.

Insbesondere der **Sicherheitsvorteil** von Sprachbedienung wurde in verschiedenen Studien thematisiert (vgl. Tsimhoni et al., 2004; Shneidermann, 2002). So konnten Young et al. (2003) zeigen, dass Navigationssysteme mit Spracherkennung ergonomischer und sicherer sind als Systeme, die visuell-manuellen Input benötigen. Dies wird von Vollrath und Totzke (2000) gestützt, die belegen, dass sich Sprachdialogsysteme im Vergleich zu visuell-haptischen Bediensystemen nicht negativ auf die Fähigkeit des Nutzers zur Längs- und Querführung des Fahrzeugs auswirken. Auch Maciej und Vollrath (2009) konnten in einer Fahrsimulator-Studie belegen, dass die Bedienung verschiedener Infotainmentfunktionen (Medienauswahl, Telefonbedienung, Adresseingabe) mit einem Sprachdialogsystem zu einer geringeren Ablenkung führt als die manuelle Bedienung dieser Funktionen. Bei der manuellen Bedienung konnte beobachtet werden, dass die Versuchspersonen bis zu 40% der Fahrtzeit von der Straße wegschauten. Obwohl bei der Interaktion mit dem SDS die Blickabwendung nicht völlig reduziert werden konnte, zeigt sich gerade im Bereich der Fahrleistung der positive Effekt der Spracheingabe. So führte eine Aufgabenerfüllung per Sprache zu einer geringeren Spurabweichung, weniger Korrekturen und kürzeren Reaktionszeiten. Durch die sprachliche Auswahl von Medientiteln konnte eine Reduktion der Spurabweichung von 60% erreicht werden. Bei allen manuellen Aufgaben konnte indes eine starke Beeinträchtigung identifiziert werden; so hielten die Fahrer die Spur schlechter und reagierten auf die Aufforderung zum Spurwechsel langsamer. Dass eine grundlegende kognitive Ablenkung auch bei der sprachlichen Interaktion bestehen bleibt, kann an dem Anstieg der mentalen Belastung verglichen zur Referenzfahrt beobachtet werden. Die zitierten Studien weisen eindrucksvoll nach, dass sich eine visuelle Ablenkung des Fahrers durch die Bedienung eines Infotainmentsystems problematisch auf seine Fahrleistung auswirkt und sprachliche Bedienschnittstellen diesen Effekt minimieren können. Darüber ermöglichen Sprachdialogsysteme **ergonomische Vorteile**, indem sie den Nutzer nicht dazu zwingen, in einer bestimmten Position (unnatürliche) physikalische Handlungen auszuführen (z. B. Drehen, Drücken) und damit die Hände vom Lenkrad zu nehmen.

"It does not rely on the user's hands or eyes, and it can work at a distance…" (Brennan, 1998, S. 7)

Neben der geringen visuellen und motorischen Beanspruchung, sollte man die mentale Beanspruchung in Zusammenhang mit dieser Technologie jedoch nicht außer Acht lassen. Auch wenn sprachbedienbare Schnittstellen zu einer geringeren visuell-haptischen Belastung führen und weniger Blickabwendung (engl. *eyes-off-the-road*) als grafische Schnittstellen hervorrufen, so ist ihr Einfluss auf die Fahraufgabe nicht zu unterschätzen. Je nach Komplexität des Dialogs können sie neben auditive auch mentale Beanspruchung hervorrufen, die zentrale Verarbeitungsressourcen bindet und auch als „**mind-off-the-road**"-Problem bezeichnet wird (Green, 2001; Tsimhoni et al., 2001).

Dieses Phänomen kann auch beobachtet werden, wenn Menschen sich mit den anderen Insassen des Fahrzeugs unterhalten. Im Rahmen von zwischenmenschlichen Dialogen konnte Vollrath (2007) nachweisen, dass der Effekt der mentalen Belastung durch die Art der Gespräche beeinflusst wird. Handelt es sich um normale, sachliche Gespräche wirken sie sich sehr viel weniger nachteilig auf die primäre Fahraufgabe aus, als komplexe emotionale Gespräche. Eine Meta-Analyse von über 30 Studien zur Auswirkungen von Handynutzung während der Fahrt (Caird et al., 2008) konnte belegen, dass neben der händigen Nutzung eines Mobiltelefons auch das freihändige Telefonieren negativen Einfluss auf die Fahrleistung hatte.

Bezogen auf die Mensch-Maschine-Kommunikation im Fahrzeug kann die Interaktion mit einem SDS somit lediglich als **bedingt sichere Interaktionsmodalität** bezeichnet werden, da sie auch zu dem mind-off-the-road-Phänomen führen kann. Die Sprachbedienung stellt somit kein Allheilmittel dar, das die potentielle Ablenkung einer Zweitaufgabe völlig eliminieren kann. Lee et al. (2001) untersuchten, wie sehr sprachbedienbare Emailapplikationen Einfluss auf die Reaktionszeit der Nutzer und deren mentale Beanspruchung nehmen. Verglichen zu einer Referenzfahrt konnte durch die Sprachbedienung ein Anstieg der kognitiven Beanspruchung festgestellt und ein Anstieg der Reaktionszeit um 30% beobachtet werden. So reagierten Fahrer, die ein komplexes Emailsystem per Sprache bedienten auf ein periodisch bremsendes Vorderfahrzeug signifikant langsamer. Owens et al. (2010) verglichen die manuelle Bedienung eines Mobiltelefons und eines MP3-Players mit einer sprachlichen Bedienung über das Ford SYNC® System. Sie fanden heraus, dass beide Bedienungen verglichen zu der Referenzfahrt in einer erhöhten mentalen Belastung resultierten.

An dieser Stelle kann die Gültigkeit des Modells der multiplen Verarbeitungsressourcen als eingeschränkt betrachtet werden. Sobald die Sprachbedienung ein zu komplexes Niveau erreicht und der Dialog nicht mehr mit einem normalen zwischenmenschlichen Gespräch vergleichbar ist, können Ablenkungseffekte auf die Fahraufgabe nachgewiesen werden. Übersteigt die Anstrengung für die Zweitaufgabe also ein gewisses Niveau, so sind auch bei unterschiedlichen Ressourcen Interferenzen zu beobachten. Einige Autoren diskutieren deshalb, dass eine kapazitätsbeschränkte Single-Ressourcen-Theorie, den Einfluss der Sprachbedienung auf die Fahr-

sicherheit besser beschreiben kann (Moray, 1999). Wichtiger als die spezifische Klassifikation der Aufmerksamkeitsressourcen erscheint jedoch, ob die Aufgaben **kontrolliert oder automatisch** ablaufen können. Denn bei jedem gleichzeitigen Ausführen von verschiedenen Aufgaben hängt die Höhe der Interferenz auch stark von der Erfahrung und Übung des Nutzers bei Ausführung der Handlung ab (Wickens, 2002; Kahnemann, 1973). Es kann davon ausgegangen werden, dass normale, zwischenmenschliche Dialoge sehr weit in Richtung automatische Prozesse eingeordnet werden können und damit nur gering mit der Primäraufgabe interferieren. Werden die Gespräche jedoch komplexer, im Fall der Sprachdialogsysteme vielleicht auch undurchsichtiger, so wandern sie auf dem Kontinuum immer weiter in Richtung kontrollierter Prozesse und benötigen mehr Kapazität der zentralen Aufmerksamkeitsressource.

Es lässt sich demnach schlussfolgern, dass Nutzerschnittstellen, die Spracheingabe ermöglichen, nur unter bestimmten Bedingungen das Potential bieten, die Ablenkung bei der Bedienung des Infotainmentsystems zu reduzieren. Vor dem Hintergrund, dass eine besondere Herausforderung im Bereich der automotiven SDS darin liegt, dass sie als Nebenaufgabe zur primären Fahraufgabe ausgeführt wird und damit nicht die volle Aufmerksamkeit des Fahrers binden darf, muss sie so gestaltet sein, dass sie eine gewisse Komplexität nicht überschreitet. Der Gestaltung der Benutzungsschnittstelle und dem Dialogablauf kommt somit eine gewichtige Rolle zu, um das SDS als sichere Modalität zu etablieren. Dabei muss gewährleistet sein, dass die Dialoge ein mentales Beanspruchungslevel, welches typisch für eine normale zwischenmenschliche Kommunikation ist, nicht überschreiten und praktisch nebenbei geführt werden können. Das heißt, der Sprachdialog sollte möglichst verständlich und wenig frustrierend gestaltet sein, um den Fahrer sinnvoll während der Fahrt unterstützen zu können. Bei Gegebenheit dieser Bedingung bietet sich Sprache als eine optimale Mensch-Maschine-Schnittstelle zur Erfüllung tertiärer Aufgaben im Fahrtkontext an.

Bedingt durch gestiegene Nutzerbedürfnisse, steigen seit Jahren die Anzahl und der Umfang der **Funktionen** im Fahrzeug an (Hassel, 2006). Die MTV Jugendstudie (2010) belegte, dass junge Deutsche auf die Nutzung von Kommunikationsmitteln wie Smartphones, während der Fahrt nicht verzichten möchten. Bratzel (2011) stützt diesen Befund und berichtet, dass ein Drittel der jungen Deutschen gern die Möglichkeit hätte, in ihrem Fahrzeug Emails zu empfangen und zu bearbeiten. Weitere Erhebungen zeigen bereits, dass 40% der Fahrer, die ein Smartphone besitzen, es während der Fahrt trotz der hohen Sicherheitsrisiken zum Schreiben von Mails oder Textnachrichten nutzen (Schreiner, 2011). Die Integration erster Online-Dienste, die Anbindung von digitalen Komponenten (z. B. digitales Radio) und die Einbindung alltäglicher Kommunikationsformen (SMS, Email oder Facebook) halten daher in die ersten automotiven Infotainmentsysteme Einzug und ermöglichen dem Nutzer damit den Zugang zu umfangreichen Daten. Auch wenn damit die nutzerseitigen Kommunikationswünsche erfüllt werden, darf die Primäraufgabe des Fahrens nicht beeinträchtigt werden (Totzke, 2001).

Insbesondere die Sprachbedienung bietet hier das Potential, jene Funktionen mit einer sicheren Schnittstelle während der Fahrt bedienbar zu gestalten. Die sichere Bedienbarkeit komplexer Datenmengen kann nur gewährleistet werden, wenn Sprache als Tastaturersatz und damit als Alternative zur visuell-haptischen Eingabe angeboten wird. Stellt man sich beispielsweise vor, dass Nutzern ein sprachlicher Zugriff auf die Medieninhalte ihres Smartphones ermöglicht wird, so würde die Sprachbedienung einen komfortablen und teilweise sogar schnelleren Zugriff auf einzelne Funktionen als eine haptische Bedienung erlauben (Hamerich, 2009). Denn alle Medientitel wären per Sprache direkt erreichbar und bedürften keines mühsamen Durchlaufens hierarchischer Menüs oder langer Listen. Neben diesem direkten und menü-freien Ansprechen von Funktionen, bieten SDS auch die Möglichkeit **unscharfe Such-Anfragen** zu äußern. Ungenaue Anfragen, wie „Spiel mal was aus den 80zigern!“ sind denkbar, da aus dieser Anfrage systemseitig das Genre 80ziger Jahre als Suchfilter extrahiert werden kann.

Neben einer sicheren und effizienten Interaktionsmodalität handelt es sich bei Sprache um eine der natürlichsten Eingabeformen. Das folgende Kapitel adressiert den intuitiven Charakter der Sprachbedienung.

2.3.2 Sprache als bedingt intuitive Bedienmodalität

Grundsätzlich handelt es sich bei jeder Art von Mensch-Maschine-Interaktion um einen Dialog, der einen schrittweisen und wechselseitigen Austausch von Informationen und Intentionen darstellt. Bei nicht-sprachlichen Nutzerschnittstellen bleibt dieser **Dialogbegriff**, wie er in Kapitel 2.2 eingeführt wurde, jedoch metaphorischer Natur (Harris, 2005). Nur Sprachdialogsysteme bedienen sich der zwischenmenschlich wichtigsten Modalität der Kommunikation. Sie ist dem Nutzer unmittelbar zugänglich und muss daher nicht explizit erlernt werden (Wahlster, 2000). Da jeder Nutzer ein Experte in der alltäglichen Verwendung von Sprache ist, entfallen längere Anlernzeiten zur Schnittstellenbedienung. Der Nutzer muss seine Ziele nicht in Mausklicks oder Menüstrukturen übersetzen, sondern kann im besten Fall sagen was er möchte.

"Voice User Interfaces complement and at times replace graphical user interfaces, freeing people from the constraints of ‘WIMP’ (windows, icons, menus and pointers)." (Nass & Brave, 2005, S. 3)

Auch wenn Sprache als Bedienmodalität der Schnittstelle nicht erlernt werden muss, so besitzen Nutzer andererseits keine ausgeprägten Strategien im Dialog mit maschinellen Sprachverarbeitungssystemen. Die Funktionsweisen und Dialogführung der SDS unterscheiden sich teilweise so stark von den gewohnten zwischenmenschlichen Gesprächen, dass zwar die Interaktionsmodalität gleich ist, der Transfer bekannter Strategien aber nicht immer möglich ist. Den Nutzern sind der Ablauf der Spracherkennung und seine Grenzen nur selten bekannt. Oft wissen Nutzer nicht, was sie sagen können oder sind sich über den Status des Systems nicht bewusst.

Ohne das Handbuch zu studieren und einen Blick auf die möglichen Kommandos zu werfen, können Aufgaben nur mit großem Aufwand erledigt werden. So sind das mangelnde Wissen über Kommandos oder Einstiegspunkte, sowie über die Dialogabläufe häufige Bedienhürden. Ein Einstieg in das System erfordert damit, trotz der zuvor gelobten Intuitivität der Schnittstelle eine anfängliche Lernphase (Salmen, 2002), da der Nutzer sich erst mit den Möglichkeiten und Funktionen eines SDS vertraut machen muss.

Eine **grafische Bedienschnittstelle** hat hier gewisse Vorteile und ermöglicht ein vergleichsweise schnelles Erfassen der angebotenen Inhalte und Funktionen. Harris (2005) bezeichnet dies als das Tastaturmodell (engl. *keypad model*), in dem die Maschine dem Nutzer alle Optionen zur Bedienung klar aufzeigt (siehe Abbildung 6). Damit wird der Umfang der Nutzereingaben zwar limitiert, aber stets transparent vermittelt. Der gegensätzliche Pol, das Konversationsmodell (engl. *conversational model*) ermöglicht mehr Freiheiten hinsichtlich des Inputs, damit einhergehend aber auch eine höhere Unsicherheit, welche Funktionen sprachbedienbar und welche Worte im Vokabular enthalten sind.

Abbildung 6: Keypad vs. Conversational Model

Während die schriftliche Darbietung auf direkten oder indirekten Bedienschnittstellen demnach alle zur Verfügung stehenden Optionen anzeigt und somit eine geringe nutzerseitige Unsicherheit über die Systemkompetenzen besteht, vermittelt die Modalität der Sprache keine transparenten Bediengrenzen. Der speak-what-you-see-Ansatz, bei dem mögliche Kommandos je nach Kontext variierend angezeigt werden, kann als ein erster Schritt betrachtet werden, das conversational model der Sprachbedienung dem keypad model ein wenig näher zu bringen und damit nutzerseitige Unsicherheit zu reduzieren.

Auch zur Darbietung umfangreicher akustischer Informationen ist Sprache als Ausgabemodalität nur bedingt geeignet, da die Speicherkapazität oder **Gedächtnisspanne** des menschlichen Kurzzeitgedächtnisses (KZG) als begrenzt gelten kann. Miller (1956) untersuchte erstmals wie

viele Informationseinheiten ein durchschnittlicher Proband für kurze Zeit im Gedächtnis behalten kann. Die Studie ergab, dass im Mittel 7(±2) Informationseinheiten (engl. *Chunks*) unter Laborbedingungen memoriert werden konnten (Miller, 1956; Zimbardo et al., 1999). Da im Fahrtkontext nicht die volle Aufmerksamkeit für das Sprachdialogsystem aufgewendet werden kann, kann die Gedächtnisspanne für die Sprachausgaben als noch stärker begrenzt angesehen werden. Hof (2007) schätzt sie auf 4(±2) Informationseinheiten.

Der Nachteil der akustischen Vermittlung von Information liegt darüber hinaus in ihrem Mangel an Beständigkeit. Sprache ist temporal und damit ohne persistenten Charakter. Bei der Sprachausgabe ist der Nutzer demnach dazu angehalten, die Informationseinheiten zu memorieren. Ebenso kann er den Zeitpunkt der Informationsaufnahme nicht so leicht kontrollieren. Dieser Mangel an Langfristigkeit und wird auch häufig als **Persistenz-Problem** bezeichnet (Morgan et al. 2001). Die Beständigkeit der visuellen Darstellung ist der sprachlichen an dieser Stelle überlegen. Bei einer visuellen Darbietung kann ohne jegliche Memorierungsleistung auf einzelne Informationen leichter zurückgegriffen werden. Sprache als flüchtige Modalität ist an dieser Stelle im Nachteil.

Damit kann Sprache nur unter bestimmten Umständen als sichere und intuitive Bedienmodalität gelten. Im Folgenden Kapitel soll darauf eingegangen werden, dass auch die nutzerseitige Akzeptanz nur bedingt gegeben ist.

2.3.3 Akzeptanz von Sprachbedienung

Kapitel 2.3.1 konnte aufzeigen, dass die Implementation von Sprachbedienung im Fahrzeug mit **Sicherheitsaspekten** begründet werden kann. Auch Nutzer sind sich der kognitiven Entlastung durch Sprachbedienung durchaus bewusst und bevorzugten in der Studie von Noszko und Zimmer (2002) die Eingabe per Sprache vor der manuellen Eingabe. Dafür nennen sie vor allem sicherheitsrelevante Gründe, wie eine geringere visuelle und geringere motorische Beanspruchung. Eine weitere Untersuchung stützt diesen Befund und belegt, dass rund 75% der Probanden das SDS als erste Eingabemodalität wählten, um eine Aufgabenstellung während der Fahrt zu lösen (Kainer, 2007). Auch hier begründeten die Probanden ihre Entscheidung damit, dass diese Bedienform am wenigsten von der Primäraufgabe ablenkt.

Allerdings scheint dies vermittelt über die Priorisierung der jeweiligen Primär- bzw. Sekundäraufgabe und die Effizienz der Bedienschnittstelle. So konnte Brumby et al. (2011) nachweisen, dass Versuchspersonen bei einer priorisierten Zweitaufgabe die visuelle (und schnellere) Interaktion der sprachlichen Modalität vorzogen, auch wenn dies zu Einbußen in ihrer Fahrleistung führte. Waren die Versuchspersonen allerdings dazu angehalten die Primäraufgabe zu priorisieren, so wählten sie mehrfach die Sprachbedienung.

Dieser Befund legt nahe, dass Nutzer, auch wenn sie sich dem Potential der Sprachbedienung bewusst sind, diese nicht immer als Interaktionsmodalität präferieren. Die **Interaktions**dauer scheint dabei eine bedeutende Rolle zu spielen. So ergeben experimentelle Untersuchungen von Wechsung et al. (2010), dass die Nutzungswahrscheinlichkeit einer Modalität davon abhängt, wie viele Interaktionsschritte durchzuführen sind. Die Sprachbedienung wurde als Eingabemodalität nur dann präferiert, wenn sie die Lösung einer Aufgabe mit weniger Interaktionsschritten erlaubten. Auch Cameron (2000) belegt, dass die Benutzung des Sprachdialogsystems nur dann erfolgt, wenn die Verwendung schneller ist als andere Modalitäten oder andere Modalitäten aufgrund von Rahmenbedingungen ausscheiden.

Die Befundlage verdeutlicht, dass Nutzer nur dann eine hohe Akzeptanz gegenüber den SDS zeigen, wenn darüber ein erheblicher Anteil an Interaktionsschritten eingespart werden kann. Aus Erfahrungen mit aktuellen Sprachdialogsystemen und empirischen Untersuchung (z. B. Sodnik et al., 2008) ist allerdings bekannt, dass die Aufgabenerfüllung mit Sprachbedienschnittstellen häufig mehr Zeit in Anspruch nimmt, als die Bedienung über traditionelle grafisch-haptische Bedienschnittstellen.

Auch den Vergleich zu menschlichen Gesprächspartnern verlieren SDS bezüglich der Effizienz. So konnte Porzel (2006) im Rahmen von Wizard-of-Oz-Experimenten[6] nachweisen, dass zwischenmenschliche Sprachdialoge beinah doppelt so effizient ablaufen, wie Dialoge zwischen Menschen und Maschinen.

Die Akzeptanz der Bedienschnittstelle wird ebenfalls durch die **Zuverlässigkeit** der Spracherkennung beeinflusst. Wie bereits in Kapitel 2.2.3 beschrieben, handelt es sich bei der Sprachbedienung verglichen zu anderen Modalitäten um eine recht fehleranfällige Schnittstelle. Obwohl innerhalb von Whitebox-Tests[7] eine Worterkennungsrate von durchschnittlich 95% erreicht werden kann, bedeutet dies immer noch, dass jede zwanzigste Nutzereingabe in einer Fehlerkennung (Systemfehler) resultiert. Trotz enormer technologischer Entwicklungsschritte ist auch die

---

[6] In „Wizard of Oz“-Experimenten (WoZ) wird die Funktionalität des zu testenden Systems von einem Menschen simuliert, während der Proband glaubt, mit einer Maschine zu interagieren. Eine Übersichtsarbeit zu WoZ-Studien liefern Dahlbäck et al. (1993).

[7] Whitebox-Tests bezeichnen eine Testmethode bei der die Nutzer mit Kenntnissen über die Kompetenzen (Wortschatz) des zu testenden Systems ausgestattet werden.

Reliabilität eines SDS noch nicht mit den visuell-haptischen Alternativmodalitäten oder gar einem menschlichen Gesprächspartner vergleichbar.

Die geringe Zuverlässigkeit kann zweierlei Auswirkungen haben. Einerseits führt sie zu einer erhöhten mentalen Belastung, die sich wiederrum negativ auf die Fahrleistung der Nutzer auswirken kann (Kun et al., 2007; Hof, 2007). Andererseits wirken sich schlechte Erfahrungen negativ auf die Nutzungshäufigkeit der Schnittstelle aus und führen zu Vorbehalten seitens der Nutzer, weshalb SDS bisher eher für Frustration und Fehlerkennung bekannt geworden sind, als für Effektivität und Joy-of-use (Peissner et al., 2004). Langzeituntersuchungen zeigten, dass Nutzer eine sehr geringe Akzeptanz bei Fehlerkennungen zeigen. Nach nur zwei nicht erfolgreichen sprachlichen Interaktionen mit dem Infotainmentsystem wird auf die Benutzung des SDS verzichtet (Opfer et. al, 2007).

Die geringere Zufriedenheit der Nutzer liegt neben den Problemen der Spracherkennung auch an den unrealistischen **Erwartungen** an das Sprachverstehen eines Systems. Neben vereinzelt medial generierten Erwartungshaltungen aus Science Fiction Filmen, in denen sprachverarbeitende Systeme mit einem hohen Kompetenzspektrum gezeigt werden (z. B. KITT; Fenady et al., 1982), vergleichen Nutzer SDS auch zu menschlichen Gesprächspartnern. Dabei ist der Wortschatz aktueller SDS um ein vielfaches kleiner, als der eines durchschnittlichen Menschen. Auch sind schwere Fehlerkennungen, die die Erreichung des Interaktionsziels gefährden, nur selten in zwischenmenschlichen Gesprächen beobachtbar.

“Consequently, misunderstandings, caused by, for instance, recognition errors, which are very frequent in HCI, are very rare in HHC.” (Fischer, 1999, S. 4).

Betrachtet man allerdings den Zusammenhang der Zuverlässigkeit der Spracherkennung mit der nutzerseitigen Akzeptanz, so zeigt sich ein überraschendes Ergebnis. Die verbliebenen technischen Restriktionen der Spracherkenner können nicht als alleinige Ursache für die geringe Akzeptanz gelten. So konnte Peissner (2002) nachweisen, dass die Bedienqualität von SDS im Durchschnitt nur zu etwa 15-30% von der Leistung des Spracherkenners abhängt. Demnach scheint es für die Verbesserung der Gebrauchstauglichkeit von SDS nicht essentiell zu sein, die Erkennungsgüte um wenige Prozentpunkte zu verbessern, sondern vor allem die Dialoggestaltung so nutzerfreundlich wie möglich zu gestalten.

In dem vorangegangenen Kapiteln konnte verdeutlicht werden, dass Sprachbedienung im Fahrtkontext verglichen zu herkömmlichen Aktions- und Wahrnehmungsmodalitäten die manuell-visuelle Ablenkung reduziert und damit geringere Interferenzen mit der Fahraufgabe zeigt. Dennoch stellt die Interaktion mit SDS kein Allheilmittel dar, um die Ablenkung tertiärer Aufgaben gänzlich zu reduzieren. Bezüglich der Effizienz und Effektivität verlieren aktuelle SDS nicht nur

den Vergleich zu menschlichen Gesprächspartnern, sondern auch zu alternativen Bedienmodalitäten. Das führt dazu, dass Menschen SDS im Fahrzeug weniger nutzen, als es ihr Potential vermuten lässt. Ein gut gestalteter Dialog ist zwingend erforderlich, um die Akzeptanz zu steigern und damit die Aufmerksamkeitsabwendung von der primären Fahraufgabe weiter zu reduzieren. Um Dialoge adäquat zu designen, soll sich in der vorliegenden Abhandlung am menschlichen Kommunikationsverhalten orientiert werden. Im Folgenden Kapitel soll aufgezeigt werden, dass Interferenzen entstehen, wenn erlernte Dialogstrukturen in Mensch-Maschine-Dialogen keine Anwendung finden können.

2.3.4 Interferenzen durch erlernte Dialogprinzipien

Verschiedene Beobachtungen können belegen, dass Nutzer zwischenmenschliche Kommunikationsprinzipien auch im Dialog mit sprachverarbeitenden Maschinen anwenden. Einige Autoren postulieren daher, dass die Verarbeitungsmechanismen der Sprachproduktion und des Sprachverständnis bei beiden Formen der Interaktionen gleich ablaufen (vgl. Nass & Brave, 2005). Sie gehen davon aus, dass das bewusste Wissen, dass Sprache eine nicht-humane Quelle haben kann, nicht ausreichend ist, um bestehende Dialogverhaltensweisen oder Interpretationsstrategien zu verwerfen. So konnten Nass und Brave (2005) nachweisen, dass eine Aktivierung sozialer Reaktionen (z. B. geschlechtsspezifischer Stereotypen) über eine Computerstimme genauso möglich ist, wie über eine menschliche Stimme.

"Listeners and talkers cannot suppress their natural responses to speech, regardless of source […] applying the same rules and shortcuts that they use when interacting with people" (Nass & Brave, 2005, S. 4).

Dabei lässt sich die Befundlage nur bedingt damit erklären, dass Menschen Maschinen als soziale Akteure akzeptieren. Vielmehr fallen sie in bestehende, automatisierte Verhaltensmuster zurück, die sich in Jahrhunderten von sozialer und kognitiver Evolution von Sprache entwickelt haben. Ohne sich dem Transfer bewusst zu sein, übertragen Menschen diese erlernten Kommunikationsstrategien auf die Konversation mit Maschinen.

So bringen sie beispielsweise die Erwartungshaltung mit in den Dialog, dass dieselben **Reparaturstrategien**, die zwischenmenschlich die Erfolgswahrscheinlichkeit einer Kommunikation erhöhen, auch den Dialog mit einer Maschine effektiver gestalten können. Erneut sei auf die Phänomen des Lombard-Effekts und der Hyperartikulation aus Kapitel 2.2.3 verwiesen. Während die Tonhöhenanpassung und die Überbetonung in zwischenmenschlichen Dialogen zu einem erhöhten gegenseitigen Verständnis führen, kann der gegenteilige Effekt bei der Anwendung dieser Strategien bei maschineller Spracherkennung beobachtet werden.

Weiterhin gehen Nutzer intuitiv davon aus, Sprachdialogsysteme **unterbrechen** zu können und tätigen Spracheingaben, während das System selbst spricht oder verarbeitet. Während bei menschlichen Dialogpartnern die hochdynamischen und überlappenden Sprachäußerungen die Effizienz der Kommunikation steigern können, führen sie gegenüber SDS in den meisten Fällen zu einer Fehlerkennungen (vgl. Kapitel 2.2.3). Der folgende beispielhafte Dialog, publiziert von IBM (2001, S. 165), zeigt, dass Nutzer erwarten, systemseitige Sprachausgaben unterbrechen zu können und diesen automatisierten Prozess auch nur schwer unterdrücken können.

System	Do you want another transaction? <Be ep>
Nutzer	Yes.
System	Remember to wait for the tone. Do you want another transaction?
Nutzer	Yes.
System	<Beep>

Diese Gegensätze zeigen an, dass die Missachtung der Prinzipien der zwischenmenschlichen Kommunikation bei der Gestaltung von SDS Fehlbedienungen und Verwirrung hervorrufen kann.

In Anlehnung an Kapitel 2.3.1 kann davon ausgegangen werden, dass normale zwischenmenschliche Gespräche ohne eine große **mentale Beanspruchung** geführt werden können. Weicht der Dialog mit dem SDS zu weit von diesem Pol der automatischen Prozesse ab, so kann ein Anstieg in der Ablenkung durch die Sprachbedienung beobachtet werden, da Sprecher nur mit großem Aufwand ihre natürlichen Reaktionen auf Sprache unterdrücken können. Die Anpassung des Dialogdesigns an bestehende zwischenmenschliche Kommunikationsstrategien ist also auch relevant, um Ressourcen beim Nutzer zu sparen. Die Anwender sind derart an die Regeln des zwischenmenschlichen Austauschs gewöhnt, dass ein Umdenken nur unter großer mentaler Beanspruchung möglich ist. Als Beispiele für diese Belastungseffekte lassen sich Konversation über Funkgeräte oder Gegensprechanlagen nennen. Kommunikationshürden, die die Dialogstruktur betreffen, wie das nur einseitig mögliche Sprechen über diese Geräte, bedürfen eines gewissen Trainings.

Abseits der Orientierung am menschlichen Dialog, postulieren einige Autoren die Entwicklung einer bestimmten Kunstsprache (vgl. Schiel, 2006). Die Befundlage zeigt jedoch, dass ein **Training** von Nutzern hin zur Verwendung einer Maschinensprache nicht empfehlenswert ist (vgl. Fischer, 1999). Dieses führt vielleicht objektiv zu besserer Spracherkennung und Effizienz, aber zu höheren Belastungswerten beim Nutzer. Harris (2005) berichtet von dem prototypischen Circuit Fix-it Shoppe System von Smith und Hipp (1994) bei dem Nutzer dazu aufgefordert waren

"verbie" zu sagen, um eine Spracheingabe anzukündigen und "over" um diese zu beendeten. Diese Art von Dialogdesign ist derart kontra-intuitiv, dass die Nutzer die Kommandos häufig unterließen und damit Erkennungsfehler provozieren (ebd.). Da die Häufigkeit der Interaktionen mit SDS verglichen zu zwischenmenschlichen Interaktionen, als gering eingestuft werden kann, werden Nutzer auch bei intensivem Training diese Fehler weiterhin begehen.

Es kann geschlussfolgert werden, dass die nutzerseitigen Erwartungen an die Interaktion durch die menschliche Kommunikation geprägt sind und die Missachtung des sozialen Aspekts der Kommunikation zu Irritationen und Fehlern führt. Durch die Möglichkeit zur Anwendung zwischenmenschlicher Kommunikationsstrategien soll erreicht werden, dass das mentale Beanspruchungsniveau einer normalen zwischenmenschlichen Kommunikation nicht überschritten wird.

2.4 Fazit

Im Fokus dieser Arbeit steht die Verbesserung der sprachlichen Schnittstellen, um die sichere und angenehme Bedienung von tertiären Infotainmentfunktionen im Fahrzeug zu gewährleisten. Vor diesem Hintergrund wurden zunächst allgemeine Prinzipien der Mensch-Maschine-Interaktionen dargestellt und das Konstrukt der Gebrauchstauglichkeit eingeführt.

Sprachliche Schnittstellen bieten den Vorteil, eine hand- und augenfreie Interaktion zu ermöglichen und damit die Ressourcen, die durch die Fahraufgabe bereits gebunden sind, nicht zusätzlich zu belasten. Darüber hinaus ist Sprache als Bedienmodalität dem Nutzer intuitiv zugänglich und muss nicht explizit erlernt werden. Aufgrund jener Vorzüge haben sich SDS im Fahrzeug als integraler Bestandteil der Bedienkonzepte etabliert. Allerdings verdeutlicht die Beschreibung typischer Dialogabläufe und Erkennungsfehler, dass die Leistungsfähigkeit aktueller automotiver SDS noch Optimierungspotential bietet. Dem großen Potential von Sprachdialogsystemen im Fahrzeugkontext stehen Probleme gegenüber, die rein technisch (noch) nicht gelöst werden können.

Der Ansatz der Arbeit ist daher die Frage, ob man durch eine nutzerzentrierte Gestaltung des Dialogdesigns eine verbesserte Akzeptanz erreichen kann. In diesem Kontext wurde argumentiert, dass es sinnvoll erscheint, sich bei der Gestaltung von SDS an den Richtlinien der zwischenmenschlichen Kommunikation zu orientieren. Durch die Möglichkeit, Prinzipien der zwischenmenschlichen Interaktion auch auf Dialoge mit Maschinen anzuwenden sollen Bedienfehler und das mentale Beanspruchungslevel einer Interaktion mit einem automotiven SDS reduziert werden. Vor diesem Hintergrund werden im nächsten Abschnitt die Prinzipien der zwischenmenschlichen Dialogführung dargestellt.

3 Grundlagen zwischenmenschlicher Dialogführung

Um Gestaltungsempfehlungen für SDS an zwischenmenschlichen Kommunikationsstrategien anzulehnen, sollen zunächst gängige Kommunikationsmodelle vorgestellt und relevante Prozesse der zwischenmenschlichen Dialogführung eingeführt werden. Anhand der Collaborative Theory von Clark (1996) werden darauf aufbauend relevante Strategien zum Aufbau einer geteilten Wissensbasis identifiziert, die maßgeblich zum Erfolg einer Kommunikation beitragen können. Kapitel 3.2 fokussiert anschließend auf drei dieser Elemente, die anhand von zwischenmenschlichen Dialogbeispielen nähere Erläuterung finden.

3.1 Kommunikationsmodelle

Durch SDS als Untersuchungsgegenstand wird sich in der vorliegenden Arbeit bei der Analyse der Kommunikationsprozesse maßgeblich auf den gesprochen-sprachlichen Dialog beschränkt. Dieser gilt nach Pickering und Garrod (2004) als die natürlichste Form des Sprachgebrauchs. Neben dem Austausch von Informationen geht es dabei vor allem um die Dialogstrukturierung, die nötig ist, um eine Interaktion erfolgreich zu gestalten. Die Abhandlung baut auf einem Grundlagenmodell der Kommunikation auf, welches im nächsten Abschnitt dargestellt wird.

3.1.1 Sender-Empfänger-Modell

Ein stark vereinfachtes Abbild der Kommunikation stellt das sogenannte Sender-Empfänger-Modell dar (siehe Abbildung 7).

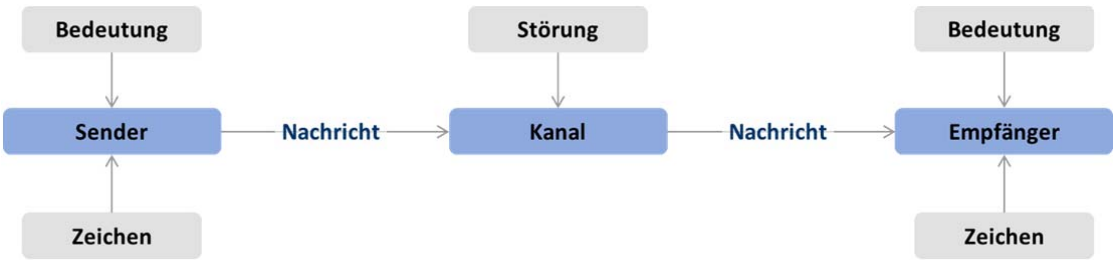

Abbildung 7: Kommunikationsmodell nach Shannon und Weaver

Das vergleichsweise strukturarme Modell geht auf die mathematische Theorie der Kommunikation von Shannon und Weaver zurück (1976; Krallmann & Ziemann, 2001). Eine *Informationsquelle* wählt eine *Nachricht* aus, die aus geschriebenen oder gesprochenen *Zeichen* bestehen kann. Der *Sender* verwandelt diese in ein *Signal*, das über einen *Kommunikationskanal* an einen *Empfänger* übertragen wird, welcher die Information dann decodiert. Durch *Störungen* können die ursprünglichen Signale beeinflusst werden (vgl. Shannon & Weaver, 1976). Bezogen auf die zwi-

schenmenschliche Kommunikation, geht man davon aus, dass jeder Verwender von Sprache einen Bedeutungsvorrat hat, welches sich als konzeptionelles Wissen (über die Welt) beschreiben lässt. Des Weiteren verfügt er über einen Zeichenvorrat, das sogenannte Sprachwissen. Der Bedeutungsvorrat ist dann über den Zeichenvorrat in eine Nachricht transformierbar. Die Bedeutung, die der Sender an den Empfänger übertragen möchte, kann durch facettenreiche Gründe gestört werden (z. B. Umgebungsgeräusche). Kann angenommen werden, dass die Zeichensequenz den Empfänger störungsfrei erreicht, so kann er über seinen Zeichenvorrat die Nachricht in ihre Bedeutungseinheiten zurück transferieren.

Das Sender-Empfänger-Modell orientiert sich stark an technischen Aspekten der Signalübertragung. Insgesamt ist das Modell zur Beschreibung sozialer Kommunikationsprozesse nur bedingt geeignet. So wird die Interdependenz des Sprechers und Empfängers nicht ausreichend dargestellt. Dass der Empfänger und die Gesprächssituation maßgeblich die Äußerung des Sprechers beeinflussen, wird nicht berücksichtigt. Es kann beispielsweise beobachtet werden, dass ein Sprecher sich für Ausdrücke entscheidet, von denen er ausgeht, dass der Empfänger sie versteht. Dies würde einen weiteren Kanal (von Empfänger zum Sender) darstellen, der in dem Kommunikationsmodell von Shannon und Weaver keine Beachtung fand. Der Pfeil der Nachrichtenübertragung ist im Originalmodell unidirektional gestaltet. Eine derart vereinfachte Strukturdarstellung wird der zwischenmenschlichen Kommunikation nicht gerecht. Ergänzend soll daher das Kommunikationsmodell von Clark (1996), die Kollaborative Theorie (engl. *Collaborative Theory*), eingeführt werden.

3.1.2 Kollaborative Theorie

Die Collaborative Theory betrachtet Kommunikation als eine gemeinschaftliche Aktion (engl. *joint actio*n) und eine Form der Kooperation. Verglichen mit autonomen Handlungen zeichnet sie sich vor allem durch die dynamische Anpassung der Verhaltensweisen an den Gesprächspartner und die Situation aus. So werden die kommunikativen Aktionen nicht nur in Bezug aufeinander ausgeführt, sondern auch hochgradig miteinander koordiniert.

Clark (ebd.) betrachtet die Konzepte des Sprachgebrauchs und der gemeinschaftlichen Aktionen als untrennbar. So können Interaktionspartner keine joint action (z. B. Tanzen) ausüben, ohne zu kommunizieren. Die meisten gemeinschaftlichen Aktionen - so auch Dialoge, werden durch Sequenzen von kleinen koordinierten Aktionen zusammengesetzt, die aufeinander aufbauen und in einer starken Abhängigkeit zueinander stehen (ebd.).

Für den Gebrauch von Sprache stellte Clark (ebd.) sechs Thesen auf, die im folgenden Erläuterung finden.

1. Sprache wird im Wesentlichen für **soziale Zwecke** verwendet.

   Sprache allein hat keine Funktion. Sie wird lediglich als ein Instrument zum Ausführen von sozialen Aktivitäten genutzt.

2. Der Gebrauch von Sprache stellt eine **kollaborative Aktion** dar.

   Jede Verwendung von Sprache benötigt mindestens zwei Agenten. Die Gesprächspartner können dabei real oder imaginär, individuelle Menschen oder Institutionen sein. Sprache, und Dialoge im Speziellen, werden als eine gemeinsame und koordinierte Handlung verstanden, bei der beide Gesprächspartner Verantwortung zur Erreichung des Interaktionsziels tragen.

3. Der Gebrauch von Sprache involviert immer die **Intention** des Sprechers und das **Verständnis** des Empfängers.

4. Der elementarste Rahmen für den Gebrauch von Sprache ist das Gespräch von Angesicht zu Angesicht (engl. ***face-to-face***).

   Sprache entstand beinah ausschließlich in gesprochen-sprachlichen Situationen. Auch Kinder lernen sprechen durch Konversationen von Angesicht zu Angesicht. Daher besitzt die face-to-face Situation eine hohe Universalität und benötigt keine Lernphase. Alle anderen Kontexte, in denen Sprache verwendet wird, werden als Derivate der face-to-face Situation gesehen.

5. Der Gebrauch von Sprache füllt häufig mehr als eine **Ebene**.

   Die einfache face-to-face Konversation verfügt häufig nur über eine Ebene der Erzählung. Jeder Gesprächsteilnehmer kann allerdings durch das Erzählen einer Geschichte oder durch das Einnehmen einer anderen Rolle weitere Ebenen der Gesprächsführung hinzufügen.

6. Zur Erforschung des Sprachgebrauchs müssen **Kognitions- und Sozialwissenschaft** kombiniert betrachtet werden.

   Die Betrachtung der Sprache lässt eine soziale und eine individuelle Perspektive zu. Einerseits können die individuellen kognitiven Prozesse eines Sprechers oder Empfängers Gegenstand der Untersuchung sein. Andererseits kann die Erforschung der sozialen Prozesse zwischen den Gesprächsteilnehmern im Mittelpunkt des Interesses stehen. Um ein allumfassendes Abbild der Realität zu erhalten, müssen jedoch beide Bereiche kombiniert betrachtet werden.

Die Ein-Ebenen Konversation, auf die sich beschränkt werden soll, wird in den folgenden beiden Absätzen hinsichtlich ihrer kognitiven und sozialen Prozesse untersucht. Zunächst sollen die prozessorientierten bzw. prozeduralen Ziele der Interaktion betrachtet werden.

3.1.2.1 Prozedurale Ziele der Kommunikation

In den meisten Fällen wird der Nutzer eines SDS die gemeinschaftliche Aktion, den Dialog, starten, um ein dominantes Ziel (z. B. Eingabe eine Navigationsziels) zu erreichen (Kap. 2.2). Neben diesen interaktionsstiftenden Motiven haben Gesprächsteilnehmer auch sogenannte interpersonelle Ziele und prozedurale Ziele.

Da nur bedingt davon ausgegangen werden kann, dass Nutzer private oder interpersonelle Ziele verfolgen, während sie mit einer sprachverarbeitenden Maschine interagieren, soll lediglich auf die prozeduralen Ziele näher eingegangen werden, die die Ansprüche der Gesprächsteilnehmer an den Prozess der Kommunikation darstellen. Sie lassen sich am ehesten durch die **Konversationsmaximen** beschreiben, die Grice (1975) für jegliche Art von Kommunikation definierte. Ähnlich wie Clark (1996) postulierte er, dass ein Dialog keine Aneinanderreihung unzusammenhängender Äußerungen sei, sondern durch Aufwand aller Gesprächsteilnehmer zu einer kooperativen Handlung erwachse. Der Aufwand lässt sich dabei in vier Kategorien gliedern, die Grice „Quantität“, „Qualität“, „Relevanz“ und „Art und Weise“ nennt und die in Tabelle 1 näher erläutert werden. Dabei liegt stets die Annahme zugrunde, dass das prozedurale Ziel einer Konversation ein maximal effektiver und effizienter Informationsaustausch sei.

Tabelle 1: Kommunikationsmaximen nach Grice (1975)

Maxime	
Quantität	Sei so informativ wie nötig. Sei nicht informativer als nötig.
Qualität	Leiste wahre Beiträge. Sage nichts, was vermutlich falsch ist. Sage nichts, wofür der Nachweis fehlt.
Relevanz	Bleibe beim Thema. Relevanz ist im Dialogverlauf veränderlich und nicht vordefiniert.
Art und Weise	Sei verständlich. Vermeide Mehrdeutigkeiten und Unklarheiten des Ausdrucks. Halte dich kurz, schweife nicht ab. Sei systematisch und geordnet.

Wenn Gesprächspartner immer ein kooperatives Verhalten zeigen und sich an diesen vier Maximen orientieren, kann ein erfolgreicher Dialog zur Erreichung eines Dialogziels gewährleistet und die Prozessziele der Gesprächspartner erreicht werden. Die Regeln stellen somit auch eine Erwartung von Gesprächsteilnehmern an den jeweiligen Dialogpartner dar.

Neben diesen prozeduralen Dialogmotiven sollen im nachfolgenden Kapitel jene kognitiven Prozesse dargestellt werden, die bei Sprecher und Empfänger während eines Gespräches ablaufen. Obwohl der Fokus auf der Verknüpfung der sprecher- und empfängerseitigen Verarbeitungsvorgängen liegt, soll zunächst ein Prozessmodell der Sprachproduktion eingeführt werden. Daran soll abschließend abgeleitet werden, welche Rolle dem Empfänger in jeder Phase der Sprachproduktion zukommt.

3.1.2.2 Kognitive Prozesse der Sprachproduktion

Laut Levelt (1989) lässt sich die Sprachproduktion in drei Hauptaktivitäten untergliedern; die Konzeptualisierung, Formulierung und Artikulierung einer Nachricht (siehe Abbildung 8).

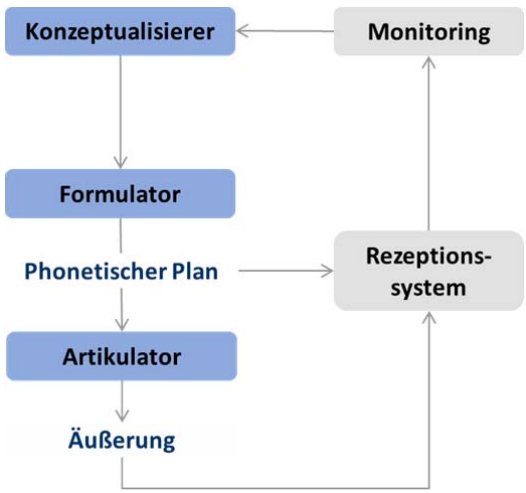

Abbildung 8: Sprachproduktion

Diesen Phasen werden unterschiedliche Prozesse zugeordnet. Der **Konzeptualisierer** dient der Generierung des Nachrichteninhalts. Dabei werden die Inhalte, die der Sprecher mitteilen möchte, in Übereinstimmung mit den dominanten und prozeduralen Konversationszielen konstruiert. In der **Formulierungsphase** wird die Bedeutung mittels syntaktischer Regeln in eine sprachliche Mitteilung enkodiert. Die **physische Exekution** (Sprechen oder Schreiben) der Kommunikationsinhalte ist Gegenstand der Artikulationsphase.

Das Modell versteht die Phasen der Sprachproduktion sowohl als sequentiell, als auch als parallel. Eine spätere Phase der Sprachproduktion kann dabei erst starten, wenn die vorangehende

Phase bereits begonnen hat. Dabei unterliegen nur die Konzeptualisierungsprozesse und die parallele Überwachung (engl. *monitoring*) der Aufmerksamkeitssteuerung und laufen kontrolliert ab. Bei bestimmten situativen Erfordernissen (z. B. fremdsprachlicher Sprachproduktion) können auch die anderen Aspekte der Sprachproduktion kontrolliert ablaufen.

Da der Gebrauch von Sprache immer die Intention des Sprechers und das Verständnis des Empfängers involviert (These 3; Clark, 1996), soll dieses Modell um Handlungen ergänzt werden, die bei der Rezeption einer Nachricht ablaufen. Auch hier lassen sich drei Phasen identifizieren. Zuerst wird der Empfänger die Nachricht wahrnehmen, dann Worte und Phrasen identifizieren, um abschließend die Bedeutung zu extrahieren. Jeder sprachproduzierenden Handlung kann somit eine korrespondierende Empfängerhandlung gegenüber gestellt werden. Zusammen bilden diese Paare die *joint actions* und verdeutlichen die Abhängigkeit der Produktions- und Verstehensprozesse (siehe Tabelle 2).

Tabelle 2: Sprachproduktionsprozesse und korrespondierende Empfangshandlungen

Sprecher	**Empfänger**
Artikulieren	Wahrnehmen
Formulieren	Identifizieren
Konzeptualisieren	Verstehen

Während eines solchen Austauschs entsteht eine geteilte Wissensbasis, die sowohl der Sprecher bei der Konzeptualisierung, als auch der Empfänger beim Verstehen einer Nachricht einbezieht. Diese wird auch als „*Common Ground*“ bezeichnet und ist Gegenstand des folgenden Kapitels.

3.1.2.3 Aufbau einer geteilten Wissensbasis

Menschen, die sich an Konversationen beteiligen, bringen grundsätzlichen einen gewissen Wissensstand, frühere Erfahrungen und Annahmen über den Empfänger in die Konversation ein. Ein Teil dieses Wissens (z. B. die gemeinsame Sprache) setzen sie auch beim Gesprächspartner voraus und gehen somit davon aus, dass dieser ausreichend Kenntnisse besitzt, die von ihnen gesendete Botschaft zu entschlüsseln.

Diese Hintergrundannahmen der Konversation werden auch als der sogenannte **Common Ground** (CG) bezeichnet und durch Stalnaker (1978) wie folgt definiert:

"Roughly speaking, the presuppositions of a speaker are the propositions whose truth he takes for granted as part of the background of the conversation. [...] Presuppositions are what is taken

by the speaker to be the common ground of the participants in the conversation, what is treated as their common knowledge or mutual knowledge." (ebd., S. 320)

Clark (1996) bezeichnet den CG als die Basis aller gemeinschaftlichen Handlungen. So ist es nach Clark nahezu unmöglich, die Sprecherintention und das empfängerseitige Verstehen miteinander zu koordinieren ohne auf den CG Bezug zu nehmen. Für die Gesprächspartner ist es notwendig, voneinander wechselseitig zu wissen, welche Informationen dem Partner verfügbar sind, um das Kommunikationsziel zu erreichen. Dabei ist der CG der Konversation aber nur selten expliziter Bestandteil der Konversation. Vielmehr kann er als ein impliziter Teil des Wissens der Gesprächspartner betrachtet werden, auf dem die Dialogführung aufgebaut wird. Die geteilte Wissensbasis hat dabei starken Einfluss auf die Planungs- oder Ausführungsphase der Sprachproduktion. So wird der Sprecher zur Vermittlung seiner Inhalte zunächst diagnostizieren, welches Wissen er bei dem Hörer voraussetzen kann und eine entsprechende Formulierung wählen. Es finden sich viele Belege, dass Sprecher sich für Formulierungen entscheiden, von denen sie ausgehen, dass sie die Wahrscheinlichkeit für eine erfolgreiche Kommunikation erhöhen. Bell (1984) bezeichnete dies als Rezipienten oder **Audience Design**.

Das im vorangegangenen Kapitel eingeführte Sprachproduktionsmodell kann somit nicht nur um die korrespondierenden Empfängerhandlungen, sondern auch um eine weitere dialogorientierte Ebene ergänzt werden (siehe Abbildung 9). Der (antizipierte) CG nimmt daher Einfluss auf die Planungsphase des Sprechakts und wird durch die Gesprächssituation geprägt.

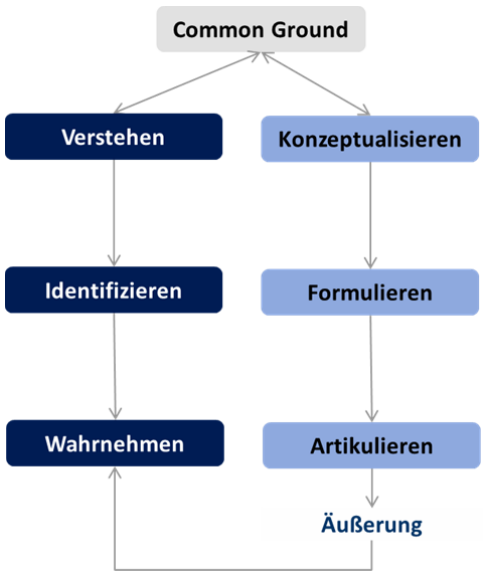

Abbildung 9: Dialogorientierung in Sprachproduktion und -verstehen

Die **Etablierung des CG** geschieht über die Akkumulation von Informationen und kann Vorannahmen bestätigen oder ändern. Somit wird jede Äußerung - oder globaler ausgedrückt - jeder

kommunikative Akt von dem CG beeinflusst und beeinflusst gleichzeitig den CG. Im Rahmen der Collaborative Theory nach Clark (1996) wird der Prozess zum Aufbau eines CG auch als **Kollaboration** bezeichnet. Eine essentielle Voraussetzung für die Bildung eines deckungsgleichen und damit gültigen CG ist, dass jede Äußerung wie intendiert verstanden wird. Dabei wird der Inhalt einer Äußerung nicht automatisch zum CG hinzugefügt, sondern erst nachdem die Gesprächsteilnehmer durch positive Rückmeldungen der Aufnahme einer Information zur geteilten Wissensbasis zugestimmt haben.

Bezugnehmend auf die Phasen der Sprachproduktion existieren nach Horton und Keysar (1996) zwei Möglichkeiten, zu welchem Zeitpunkt auf den Common Ground referenziert wird. Das **Initial Design Model** geht davon aus, dass der CG bereits beim Planen einer Äußerung in Betracht gezogen wird. Menschen werden hierbei als kooperativ interagierend aufgefasst. Konkurrierend dazu, geht das **Monitoring and Adjustment Model** von einer egozentrischen Sprachproduktionsplanung aus. Ein Einbezug des CG erfolgt erst während des Sprechaktes, wenn durch das Monitoring während der Artikulation deutlich wird, dass die rudimentäre Planung nicht ausreicht. Im Rahmen von Adjustierungsprozessen wird die Sprachproduktion dann an das partnerseitige Wissen bzw. an seine Wahrnehmungsbedingungen angepasst.

Horton und Keysar (1996) konnten nachweisen, dass insbesondere unter Zeitdruck das Monitoring and Adjustment Modell Gültigkeit besitzt und Sprecher auf einen ressourcenschonenden Plan zurückgreifen, der den CG anfänglich nicht berücksichtigt. Dies lässt sich damit begründen, dass zur Etablierung des CG der Aufbau und die Aufrechterhaltung eines komplexen Modells der Konversationssituation erforderlich sind. Der Gesprächsteilnehmer muss sowohl sein eigenes Wissen, als auch das Wissen von dem er annimmt, dass er es mit dem Partner teilt, präsent halten und seine Äußerung an das Vorwissen und den Aufmerksamkeitsfokus des Adressaten anpassen. Diese Vorgänge zum Einbezug des CG sind sehr kostenintensiv und für durchschnittliche Konversationen zumeist unnötig. Der Großteil der zwischenmenschlichen Kommunikation gelingt auch ohne den expliziten Einbezug des CG in Sprechplanung, da sich die Sprecherperspektive meist nur wenig von der gemeinsamen Perspektive unterscheidet. Auch Pickering und Garrod (2004) postulieren im Rahmen des Interactive Alignment Models (Kap. 3.2.3), dass für einfache Konversationen die Gültigkeit des Monitoring and Adjustment Models angenommen werden kann. Die explizite Bezugnahme auf den CG und damit die kontrollierte Etablierung desselbigen wird häufig als eine Abweichung vom normalen Konversationsfluss betrachtet, um Missverständnisse zu vermeiden oder zu korrigieren.

Nach Clark (1996) nutzen Sprecher im Rahmen des Initial Design Models maßgeblich zwei Arten von Hinweisen, um ihren Gesprächspartner hinsichtlich seines Wissenstandes und linguistischer Kompetenz einzuschätzen. Ein Anhaltspunkt, dem sich der Sprecher einer Äußerung bedient, ist sein eigener **Erfahrungsschatz** mit diesem bestimmten Gesprächspartner. Dieser Einbezug der

Dialoghistorie wird im Kapitel 3.2.3 noch einmal ausführlich behandelt. Als weiterer Anhaltspunkt kann die durch den Sprecher geschätzte **Kulturgemeinschaft** und linguistische Kompetenz des Empfängers gelten. Beispielsweise wird ein Sprecher eine Ortsbeschreibung daran anpassen, ob der Empfänger ein Ortskundiger oder Fremder ist, um die Erfolgswahrscheinlichkeit der Kommunikation zu erhöhen.

Insgesamt zeigen Menschen eine große Kompetenz, ihre Konzeptualisierung und Formulierung an den Empfänger anzupassen. Dies kann eindrucksvoll bei Interaktionen mit Kindern oder Ausländern beobachtet werden. Häufig bestimmt dabei die Einschätzungen der linguistischen Kompetenz des Gesprächspartners nicht nur den Inhalt einer Äußerung, sondern auch ihre Form. So sprechen Menschen zu einem Ausländer beispielsweise langsamer und verwenden einen nicht so komplexen Satzbau (engl. *foreigner talk*; nach Ferguson, 1975). Die systematischen Veränderungen der Sprachproduktion werden auch als **Register**[8] bezeichnet.

In zwischenmenschlichen Gesprächen bietet der CG nicht nur eine Repräsentation des gegenteiligen Verständnisses und steigert damit die Effektivität einer zielgerichteten Konversation, sondern kann auch die Effizienz der Interaktion steigern. Clark und Wilkes-Gibbs (1986) konnten belegen, dass sich im Verlauf einer Konversation zu einer bestimmten Aufgabenstellung die Gesprächsbeiträge deutlich verkürzen. Über die Dialoge hinweg nahm die Länge der Äußerungen ab, da sich Sprecher und Empfänger bereits auf einen CG beziehen und damit den zeitlichen Aufwand zur Gesprächsführung reduzieren konnten.

Im folgenden Kapitel sollen Prozesse eingeführt werden, die zur Koordination von Wissensständen erforderlich sind.

3.2 Grounding

Wie bereits erwähnt wurde, wird der Inhalt einer Äußerung nicht automatisch zum CG hinzugefügt, sondern die Diskursteilnehmer müssen positive Rückmeldungen dafür liefern. Das Hinzufügen von Inhalten zum CG wird auch als Grounding bezeichnet (Clark & Schaefer, 1989, Clark & Brennan, 1991). Die Konversationsteilnehmer versuchen dabei einen Zustand zu erreichen, in dem sie gegenseitig glauben, dass der Andere jeweils verstanden hat, was mit einer Äußerung gemeint wurde.

---

[8] Es besteht eine breite Befundlage zur Registertheorie und den Genren Baby oder Foreigner Talk (vgl. Ferguson, 1977).

„During the process of grounding, people exchange evidence until they reach the mutual belief that they are talking about the same thing." (Brennan, 1998, S. 10)

Als Diskursmodell des Groundings soll der psychologisch orientierte Ansatz des **Kontributionsmodells** (engl. *Contribution Models*) dargestellt werden. Theoretische Grundlage des Modells ist die These Clarks (Kapitel 3.1.2), dass es sich bei jeder Art von Kommunikation um eine kollaborative Aktion handelt. Laut Contribution Model teilen Sprecher und Empfänger die Verantwortung für die Zielführung des Dialogs (Clark & Schaefer, 1989). Als Erweiterung zu Grices Konversationsmaximen (Kap. 3.1.2.1) wird im Rahmen des Contribution Models damit die Verantwortung des Empfängers betont (ebd.). Die Gesprächspartner agieren nicht abgrenzbar als Sprecher oder Zuhörer, sondern als kooperative Aktionspartner, die beide zusammenarbeiten, um eine gegenseitige Passung von Intention und Verständnis zu erreichen. Innerhalb dieser joint action sucht der Sprecher nach verbindlichen Evidenzen über das Verständnis des Empfängers, während dieser versucht, ihm diese Bestätigungen bereit zu stellen. Da die alleinige Äußerung einer Nachricht dafür nicht ausreicht, wird jede Art von Konversation in zwei Phasen unterteilt (Clark & Brennan, 1991):

> In der **Präsentationsphase** präsentiert Sprecher A Empfänger B eine Äußerung u. Das geschieht unter der Annahme, dass Sprecher A, wenn B Evidenz (Rückmeldung) e signalisiert, glauben kann, dass B versteht, was A mit u meint.
>
> In der **Akzeptanzphase** akzeptiert Empfänger B die Äußerung u durch zeigen von Evidenz e. Empfänger B demonstriert dabei, dass er glaubt, was Sprecher A mit Äußerung u meint und nimmt an, dass A bei Registrierung von e auch annimmt, dass B verstanden hat.

In Kapitel 3.1.2.2 wurde bereits eingeführt, dass jeder sprachproduzierenden Handlung eine korrespondierende Empfangshandlung gegenüber gestellt wird. Erst zusammen bilden diese Präsentations- und Akzeptanzhandlungen die joint actions. Die **Zwei-Phasen-Struktur** der Kommunikation verdeutlicht somit, dass beidseitig die Bedeutung von Aussagen konstruiert und geprüft werden (Clark & Brennan, 1991; Brennan, 2005). Eine Äußerung zählt erst nach erfolgreicher Akzeptanzphase zu dem gemeinschaftlichen CG. Dabei verlaufen die Präsentations- und Akzeptanzphase hochgradig koordiniert. So kann der Sprecher, beispielsweise während einer Äußerung durch intonatorische Mittel wie Betonung, Pausen oder Lautstärke, Bestandteile einer Äußerung markieren, bei denen er sich nicht sicher ist und explizite Rückversicherung erbittet. Den Reaktionen des Empfängers im Rahmen der Akzeptanzphase, den so genannten **Evidenzen**, kommt damit eine bedeutende Rolle zu. Je nachdem wie gut der Empfänger seine eigene Hypothese über die Bedeutung der Äußerung einschätzt, wird er seine Evidenz anpassen und dem Sprecher damit die Möglichkeit bieten, den aktuellen Zustand der Kommunikation zu erkennen (Brennan, 2005). So kann er beispielsweise über ein Stirnrunzeln anzeigen, dass er den Beitrag

seines Gesprächspartners nur bedingt verstanden hat und den Sprecher zu einer vertiefenden Erläuterung ermuntern.

Analog zu den Sprachproduktionsphasen ergeben sich laut Brennan und Hulteen (1995, S. 144) drei mögliche **Dialogzustände** (engl. *states*), in denen sich der Empfänger befinden kann (siehe Tabelle 3). Sie sind nahezu deckungsgleich mit den 3.1.2.2 in Kapitel identifizierten Empfängerhandlungen.

Tabelle 3: Dialogzustände

0	B hat nicht gehört, dass A ein u geäußert hat.
1	B hat gehört, dass A irgendein u geäußert hat.
2	B hat u korrekt identifiziert.
3	B hat verstanden was A mit u meinte.

Für jede Äußerung, die Teil von A‘s Präsentation ist, glaubt der Empfänger B, dass er in einem der vier Zustände des Verstehens ist (Clark & Schaefer, 1989). Davon hängt ab, welchen Typ der Evidenz Empfänger B wählt. Natürlich ist es das gemeinschaftliche Ziel von Sprecher A und Empfänger B, die gegenseitige Überzeugung zu erreichen, dass B in Zustand 3 ist.

Im Rahmen der Akzeptanzphase sind durch den Empfänger je nach Dialogzustand zwei Reaktionen möglich: Einerseits kann der Empfänger B anzeigen, dass die Äußerung u verstanden wurde (Zustand = 3) oder, dass er Schwierigkeiten beim Verstehen hat (Zustand< 3). Diese Verständnisevidenzen werden auch als Feedback bezeichnet und sind im folgenden Kapitel 3.2.1 erläutert.

3.2.1 Feedback

Die im vorangegangenen Abschnitt eingeführten Evidenzen informieren den Gesprächspartner über den Prozessfortschritt des Verstehens und motivieren ihn dazu fortzufahren oder seine Äußerung anzupassen. Nach Clark (1996) sind die Empfangsbestätigungen ein wichtiger Bestandteil der zweiten Dialogspur, die neben der primären Inhaltsspur des Gesprächs, die Prozesse des Verstehens und Groundings bedient. Sie beinhaltet neben den Feedbacksignalen, die dem Sprecher Verständnisevidenzen liefern oder der Disambiguierung dienen, auch dialogstrukturierende Signale, welche Sprecherwechsel initiieren oder verzögern können. Konkret bedeutet dies für Clark und Brennan (1991), dass zwei Arten von Rückmeldungen erforderlich sind, um erfolgreich zu kommunizieren.

"…to succeed, the two of them have to coordinate both the content and process" (ebd., S. 127)

In jedem einzelnen Gesprächsbeitrag, abseits der initiierenden Äußerung, gibt es eine solche Akzeptanzkomponente, die als Verständnisbekundungen das Sprecherverhalten aktiv beeinflusst. Sie können sowohl positiv als auch negativ ausfallen. Dabei zeigt negatives Feedback an, dass Schwierigkeiten beim Verstehen bestehen. Wohingegen positives Feedback (positive Evidenzen) anzeigt, dass der Prozess erfolgreich verläuft.

Die Evidenzen für den Abschluss eines Dialogzustandes können dabei explizit oder implizit erbracht werden. Explizit sind sogenannte **Acknowledgments** (dt. *Empfangsbestätigungen*) möglich, die durch paraverbale Beiträge (z. B. *hmhm* oder *ok*), wertende Äußerungen, Wiederholungen oder Satzvervollständigungen repräsentiert werden. Definiert nach Clark und Schaefer (1989) bezeichnet der Begriff Acknowledgments somit Signale, die der Empfänger nutzt, um dem Sprecher zu verdeutlichen, dass seine Äußerung für den Zweck der Kommunikation ausreichend gut verstanden wurde und er mit der Konversation fortfahren kann. Die Acknowledgments dienen primär der Rückversicherung des Sprechers, dass sein Redebeitrag störungsfrei angekommen ist und transportieren in den seltensten Fällen neue Informationen. Sie spielen vor allem prozessbegleitend eine Rolle und organisieren die Sprecherwechsel. Demnach gibt ein Empfänger entweder Feedback während er dem Sprecher zuhört oder unmittelbar während seines Übergangs von der Hörer- in die Sprecherrolle (McTear , 2004, S. 54). Die Bedeutung von dieser Art von Acknowledgments wird deutlich, wenn man die Sprachkorpusanalyse von Traum und Heeman (1996) betrachtet. Bei der Auswertung von Problemlösedialogen fanden sie heraus, dass 51% der Sprachbeiträge mit einer expliziten Empfangsbestätigung begannen oder aus dieser bestanden.

Die prozessbegleitende Form der (paraverbalen) Gesprächsbeiträge wird als **Backchannels** bezeichnet (Sacks et al., 1974; Yngve, 1970). Backchannels stellen eine Unterkategorie der Acknowledgments dar, die nicht primär zu einem Sprecherwechsel führen (Clark & Brennan, 1991), sondern eher den Beitrag des Gesprächspartners unterstützen. Harris (2005) bezeichnet sie auch als das verbale Äquivalent des Nickens, denn sie können den Gesprächsbeitrag des Partners zwar überlappen, unterbrechen ihn dabei aber nicht. Backchannels sind somit relevante Rückmeldungen, um den Prozess des Groundings anzuzeigen (Brennan, 2005). Ohne diese Feedbacksignale sind Sprecher sich häufiger unsicher, ob sie verstanden wurden und formulieren ihren Sprechbeitrag ggf. erneut.

Als Gegenstück zu den Backchannels markieren **dialogstrukturierende Signale** den Beginn oder das Ende einer Äußerung (*hmm*, *nicht?, so*) und initiieren somit einen Sprecherwechsel. Dabei sind sie selten so explizit, wie eine "Du bist dran"-Äußerung. Wahrscheinlicher sind implizite Hinweise, wie beispielsweise das Erheben der Stimme am Ende einer Frage oder Pausen von mehr als einer Sekunde. Diese Signale werden als das Ende eines Beitrages interpretiert (Harris,

2005). Paraverbale Signale, wie *ähm* oder *uhh*, die vor Gesprächspausen geäußert werden, zeigen an, dass der Sprecher noch ein wenig Zeit zum Nachdenken braucht, bevor er seine Antwort formuliert (Brennan & Williams, 1995; Smith & Clark, 1993) und einen frühzeitigen Sprecherwechsel vermeiden möchte.

Gerade im Rahmen der mündlichen Kommunikation lassen sich neben sprachlichen Äußerungen auch **nonverbale Mittel** wie Gestik oder Mimik beobachten, die als Verständnisevidenz den Grounding-Prozess verstärken. Diese nonverbalen Handlungen können subsidiär oder substituierend zu Verbalisierungen vorgenommen werden (Salmen, 2002). Als ein Beispiel kann erneut das Stirnrunzeln genannt werden, das entweder in Kombination mit sprachlichen Äußerungen oder allein auftritt.

Auf eine weitere linguistische Form der positiven Evidenz, dem **lexikalischen Spiegeln** des Sprechers, wird im Kapitel Alignment (3.2.3) näher eingegangen. Durch die Verwendung gleicher Begrifflichkeiten kann das gegenseitige Verständnis auf Inhaltsebene ausgedrückt werden.

Die Evidenzen decken damit den gesamten Zeitraum der Sprachproduktion ab. So wird der Sprecher, während er mit der Konzeptualisierung seiner Nachricht beschäftigt ist, bereits darauf achten, ob der Empfänger bereit für die Konversation ist. Die Blickrichtung und Körperhaltung lässt beispielsweise erkennen, ob der Gesprächspartner ihm seine Aufmerksamkeit zuwendet. Während der Sprecher dann seine Nachricht formuliert, wird er durch Backchannels über das Verstehen des Empfängers informiert. Wenn dieser das Ende der Äußerung antizipiert, wird er sofort im Dialog fortfahren oder Empfangsbestätigungen zeigen. Somit zeichnen sich alle Akzeptanzevidenzen durch einen hohen Überlappungsgrad aus und tragen dabei nicht nur zur Effektivität, sondern auch zur **Effizienz** der Kommunikation bei. Dass sowohl dialogstrukturierende Signale als auch Empfangsbestätigungen produziert werden, während der Interaktionspartner noch spricht, konnten Watanuki et al. (1994) nachweisen. Im Rahmen einer Analyse zwischenmenschlicher Dialoge, die in einem Empfangsbereich eines Unternehmens beobachtet werden konnten, berichteten sie, dass 67% der Empfängerbeiträge mit den Sprecheräußerungen überlappten.

Auch positive Verstehensevidenzen werden im Sinne der Effizienz zumeist implizit dargeboten. Vor allem durch die Initiierung der nächsten relevanten Äußerung oder durch bestehende Aufmerksamkeitszuwendung, liefert der Empfänger eine positive Rückmeldung (Brennan, 2005). Stellt der Sprecher dem Empfänger beispielsweise eine Frage und antwortet dieser mit einer passenden Äußerung, so gilt allein die Passung der Antwort als ausreichend bestätigendes Feedback. Im Sinne der Ressourcenschonung verbleiben die Evidenzen implizit, solange ein Dialog erfolgreich abläuft. Grundsätzlich besteht für den Sprecher kein Grund ein Missverständnis anzunehmen, solange der Empfänger keine negative Evidenz präsentiert (Clark & Brennan, 1991). Tatsächlich fordern Sprecher aber ein Mindestmaß an Sicherheit, bevor der nächste Dia-

logschritt eingeleitet wird. Wie hoch dies ist, hängt vom Kommunikationszweck und dem entsprechenden **Grounding Criterion** ab (Brennan, 2005), welches im folgenden Kapitel 3.2.1 dargestellt wird.

3.2.2 Grounding Criterion

Je nach Interaktionskontext treten die beschriebenen Feedbacksignale bzw. Akzeptanzevidenzen in unterschiedlicher Intensität auf. So ist die von beiden Partnern geforderte Form, Stärke und Menge von Verständnisbekundungen im Rahmen der Akzeptanzphase je nach gegebenem Zweck der Kommunikation variabel. Welches Mindestmaß an Sicherheit gefordert wird, bevor der nächste Dialogschritt eingeleitet wird, definiert das sogenannte **Grounding Criterion** (GC), welches je nach Konversationskontext angepasst wird (Brennan, 2005; Clark & Wilkes-Gibbs, 1986). Wie konservativ oder liberal das Grounding Criterion gewählt wird, hängt von einer impliziten Kosten-Nutzen-Abwägung ab (Brennan, 1998). Hat ein Missverständnis nur schwer wiederrufbare Konsequenzen (z. B. unwiderrufliches Löschen gespeicherter Information) und/oder sind die aufgewendeten Ressourcen für einen Sprecherwechsel zur Generierung zusätzlicher Evidenzen gering, so wird das GC eher konservativ gewählt (Brennan, 2005). In diesem Fall werden möglichst viele Einzelbeiträge des Dialogs abgesichert.

Dieses Vorgehen wird auch als **inkrementelles Grounding** beschrieben, welches zwar einerseits eine größere Sicherheit im wechselseitigen Verständnis erzielt, gleichzeitig aber die Effizienz verringert (ebd.). Ein Beispieldialog zur Übermittlung einer Telefonnummer könnte wie folgt aussehen.

Sprecher A	Hallo, sag mal hast du die Nummer von Benjamin?	Gesprächsinitiierung
Sprecher B	Ja. 0-5-3-6-1	Präsentationsphase
Sprecher A	0-5-3-6-1	Akzeptanzphase
Sprecher B	*Hmhm* 9-1-6-5-3	Bestätigung- und Präsentationsphase
Sprecher A	9-1-6-5-3	Akzeptanzphase
Sprecher B	Ja.	Bestätigung
Sprecher A	Danke.	Gesprächsfortführung

Da die Reparaturkosten einer falsch verstandenen Telefonnummer oder Emailadresse groß sind, lässt man sich diese häufig einheitenweise noch einmal bestätigen und nimmt dafür eine geringe Interaktionseffizienz in Kauf.

Die Prozesse des Groundings werden im Allgemeinen dann als erfolgreich beschrieben, wenn ein Verständnis erreicht wurde, das für den gegebenen Kommunikationszweck ausreichend ist (Clark & Schaefer, 1989).

"…the contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for current purposes." (Clark & Schaefer, 1989, S. 262)

In diesem Zusammenhang lässt sich auch der Ansatz des **Least Collaborative Effort-Modells** (LCE) aufzeigen, der beschreibt, dass beide Kommunikationspartner motiviert sind, dass die anfallenden Aufwandskosten des Gesprächs minimal sind bzw. im ausgewogenen Verhältnis zum Nutzen stehen (Clark & Wilkes-Gibbs, 1986). Es wird demnach angestrebt, das Kommunikationsziel unter kleinstmöglichem kollaborativen Aufwand (engl. *Least Collaborative Effort*) zu erreichen.

Insbesondere die prozeduralen Ziele der Effizienz, die sich auch durch die Grice'schen Maximen der Quantität (Fasse dich kurz), aber auch der Art und Weise (Vermeide irrelevante Informationen) abbilden lassen, spiegeln sich in diesem Modell wider. Dabei versuchen Konversationspartner nach dem Prinzip der Ökonomie in der Präsentations- als auch in der Akzeptanzphase den Gesamtaufwand dynamisch zu minimieren (Clark & Brennan, 1991; Clark & Schaefer, 1989). Dies erreichen sie unter anderem durch die **Flexibilisierung des Grounding Criterions**. Ein konservatives GC, und ein damit einhergehender hoher geteilter Aufwand werden beidseitig nur dann akzeptiert, wenn der Kommunikationszweck dies rechtfertigt und ein Missverständnis negative Konsequenzen hätte. Im Sinne des LCE und der Konversationsmaximen wäre es nicht legitim, ein stark konservatives GC aufrecht zu erhalten, wenn dies der Konversationskontext nicht erfordert. In dem vorangegangen Beispieldialog wäre die Fortsetzung des inkrementellen Groundings nach der Übermittlung der Telefonnummer nicht mehr akzeptabel.

Eine frühe Untersuchung belegt, dass das GC dynamisch an den Zweck der Kommunikation angepasst wird (Wilkes-Gibbs, 1986). Dialogpartnern wurden Karten von einem Stadtzentrum gegeben, die an unterschiedlichen Stellen leere Felder aufwiesen. Primäres Konversationsziel war es, diese leeren Felder zu füllen. Eine Gruppe (hohes GC) wurde instruiert, dass sie später anhand der vollständigen Karte einem Fahrer den Weg beschreiben müssten. Die andere Gruppe (niedriges GC) sollte anhand der Karte nur eine Schätzung für die Fahrtdauer einer definierten Strecke angeben. In der Gruppe mit dem hohen GC konnten signifikant längere Dialoge beobachtet werden als in der Gruppe mit niedrigem GC.

Dabei geht es beiden Gesprächspartnern nicht ausschließlich darum, den eigenen Gesprächsaufwand zu minimieren, sondern eher um eine Reduktion des Gesamtaufwands der Kommunika-

tion. So kann es vorkommen, dass ein Partner mehr Ressourcen investiert, wenn dies den **gemeinsamen Aufwand** minimiert (Brennan & Hulteen, 1995).

Eine weitere Strategie zur Minimierung des geteilten Gesprächsaufwands stellt die linguistische Angleichung dar, die im folgenden Kapitel beschrieben wird.

3.2.3 Alignment

Eine Angleichung (engl. *Alignment*) des Verhaltens auf linguistischer Ebene soll im Folgenden auch als Alignment bezeichnet werden und meint, dass sprachliche Äußerungen auf mehreren Ebenen der vorangegangenen Äußerung angepasst werden.

Wie bereits in den vorangegangenen Kapiteln Erwähnung fand, handelt es sich bei einer Konversation um eine kollaborative Aktion, in der die jeweiligen Beiträge der Gesprächsteilnehmer nicht nur aufeinander aufbauen, sondern in einer starken Abhängigkeit zueinander stehen (Clark, 1996). Diese Interkorreliertheit der Äußerungen besteht nicht nur bezogen auf den Inhalt, sondern bezieht sich auch auf die linguistische Ausgestaltung, auf die Formulierung des Sprachbeitrags (Linell, 1998). Dialoge insgesamt beinhalten eine facettenreiche Verhaltenskonvergenz der Gesprächsteilnehmer (Branigan et al., 2010).

"Clearly convergence of both non-linguistic and linguistic behavior is robust and pervasive in dialogue." (ebd., S. 2.)

Im Rahmen der non-linguistischen Ebene lassen sich Verhaltensspiegelungen vor allem im Bereich der Gesichtsausdrücke beobachten (Bavelas et al., 1986). Die allgemeine Angleichung soll in der folgenden Abhandlung als **Konvergenz** bezeichnet werden.

Umfangreiche Forschung mit natürlichen Sprachkorpora (Tannen, 1987), Feldstudien (Giles & Powesland, 1975), quantitative Untersuchungen (Gries, 2005) und Laborexperimenten (Brennan, 1996) konnten eine solche Konvergenz aber auch auf linguistischen Ebenen nachweisen. Dabei kann eine starke Koppelung von der sprachproduzierenden Handlung und der korrespondierenden Empfangshandlung beobachtet werden.

Der **Interactive Alignment Ansatz** von Pickering und Garrod (2004) liefert einen guten Überblick über die möglichen Ebenen der linguistischen Angleichung (Abbildung 10). Die Grundannahme des Modells ist, dass in erfolgreichen Dialogen die Gesprächspartner ihr **Situationsmodell** angleichen. Ein Situationsmodell wird dabei als mehrdimensionale Repräsentation der Situation definiert, die aus den Hauptkomponenten Raum, Zeit, Intention, Kausalität und den involvierten Individuen besteht. Da es innerhalb eines Dialoges kostenintensiv ist, zwei Situationsmodelle aufrecht zu erhalten und bei nicht deckungsgleichen Modellen Missverständnisse wahrscheinli-

cher werden, versuchen die Gesprächspartner (A und B) ihre Modelle anzugleichen. Diese Angleichung kann mit der Kollaboration, der Etablierung eines Common Grounds (Kap.3.1.2.3) gleich gesetzt werden. Die Autoren (ebd.) gehen davon aus, dass eine globale Angleichung der Situationsmodelle dabei aus der lokalen Angleichung auf der Ebene der linguistischen Repräsentationen resultiert.

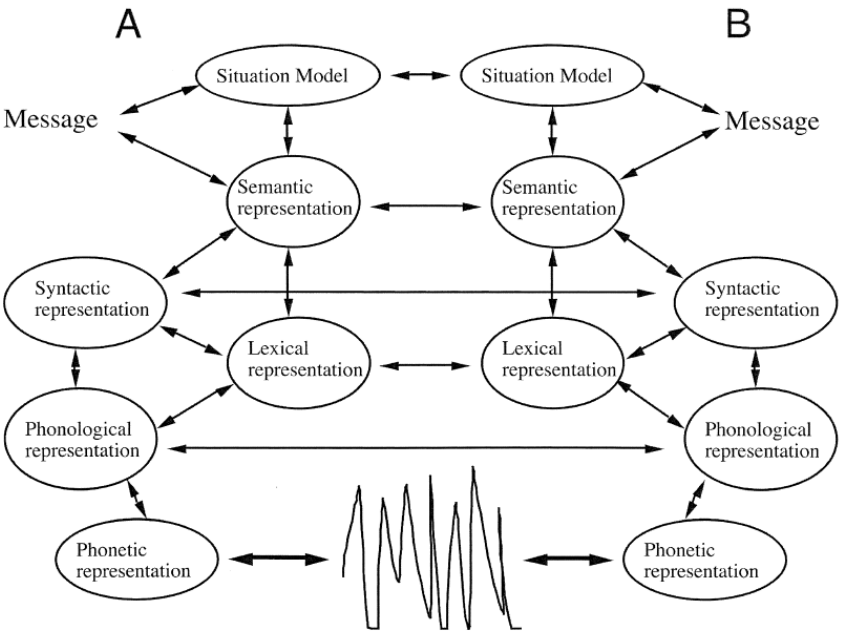

Abbildung 10: Interactive Alignment Ansatz nach Pickering und Garrod (2004, S. 176)

Eine linguistische Angleichung (gekennzeichnet durch die horizontalen Verbindungen) findet auf mehreren Ebenen statt. So konvergieren Gesprächspartner innerhalb eines Dialogs zu geteilten Phrasen und Bezeichnungen und synchronisieren mehrere Repräsentationen (Satzbau, Prosodie, Sprechrate). Eine Dialogdyade wird dann als ausgeglichen beschrieben, wenn A etwas äußert, das konsistent mit B‘s aktivierten semantischen und pragmatischen Repräsentationen des Dialogs ist und umgekehrt (Garrod & Clark, 1993). Gesprächspartner, die beide das Wort Sprachdialogsystem nutzen um auf ein sprachverarbeitendes Dialogsystem zu referenzieren, können als angeglichen (engl. *aligned*) bezeichnet werden, wobei Gesprächspartner, die nicht dieselben Begriffe wie Sprachsteuerung oder Sprachbedienung benutzen, nicht aligned sind. Weiterhin wird davon ausgegangen, dass eine Angleichung auch auf den Ebenen, die eigentlich nicht bedeutungsrelevant sind (z. B. Syntax, Betonung), eine fundamentale Rolle in der Angleichung semantischer Repräsentationen spielen. Dies liegt einer Basisannahme in ihrem Modell zugrunde, wonach eine Angleichung auf jedem linguistischen Level eine Angleichung auf den anderen Ebenen verstärkt.

Vielfältige Untersuchungen konnten belegen, dass die von dem Modell postulierten linguistischen Angleichungen auf den unterschiedlichen Sprachebenen auftreten. So konnte Pardo (2006) eine starke phonetische Ähnlichkeit bei nachgesprochenen Worten und Giles et al. (1991) eine Angleichung des Akzentes und der Sprechrate belegen. Nachweise, dass Gesprächspartner ihre syntaktischen Repräsentationen angleichen, finden sich u. a. bei Pickering und Ferreira (2008). Auch die Angleichung der grammatikalischen Ausdrucksweise zweier Gesprächspartner konnte bestätigt werden (Levelt & Kelter, 1982; Branigan et al., 2000).

Die eindrucksvollsten Ergebnisse finden sich allerdings in Untersuchungen, die die Angleichung der Wortwahl (engl. *lexical* ***Entrainment***) betreffen. In mehreren Studien konnte nachgewiesen werden, dass Gesprächspartner das gleiche Set an bezugnehmenden Äußerungen entwickeln, um auf bestimmte Objekte zu referenzieren. Die Äußerungen werden über den Dialogverlauf hinweg mit dem gleichen Partner kürzer und ähnlicher (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). Abbildung 11 verdeutlicht die Vererbungsstruktur der Begriffe Konvergenz, Alignment und Entrainment; so stellt das Entrainment eine spezifische Form des Alignments dar, welches wiederrum eine Ausprägung der allgemeinen Konvergenz ist.

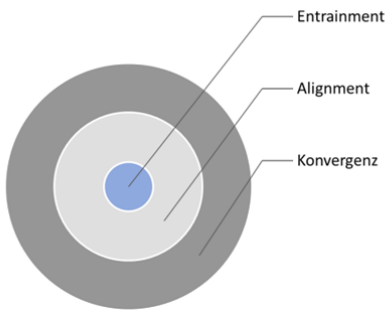

Abbildung 11: Einordnung der Begriffe Konvergenz, Alignment und Entrainment

Brennan und Clark (1996) konnten darüber hinaus beweisen, dass sich bereits synchronisierte Gesprächspartner nicht durch den Einfluss des Kontextes aus dem Gleichlauf bringen lassen. In ihrer Studie sollten Dialogpartner ein Kartenset, welches Objekte des Alltags enthielt, beschreiben und so ordnen, dass sie mit einer gleichen Kartenanordnung endeten. Waren diese Objekte aus verschiedenen semantischen Kategorien, reichte häufig der basale Begriff (Fisch), um eine Karte zu beschreiben. Waren auf den Karten allerdings Objekte aus derselben semantischen Kategorie zu sehen (Karpfen, Dorade, Nemo), so mussten die Partner in ihrer Beschreibung detaillierte Begriffe verwenden. Hatten sie sich einmal auf diese detaillierten Begriffe synchronisiert, so verwendeten sie diese auch noch in einem letzten Durchgang, bei dem eine basale Beschreibung wieder ausgereicht hätte. Somit beriefen sie sich auf die synchronisierten Konzeptualisierungen auch dann, wenn sie einfachere oder weniger informativere Begriffe hätten nutzen können. Im Rahmen dieser Untersuchungen konnte auch gezeigt werden, dass diese Effekte

partner-spezifisch sind. So verwenden Sprecher, die mit einem bestimmten Gesprächspartner interagieren, konsistent dieselbe (angeglichene) Bezeichnung. Tun sie dies nicht, so konnten Metzing und Brennan (2003) Schwierigkeiten beim Empfänger beobachten. Es entspricht also in gewisser Weise der Erwartungshaltung, dass Sprecher mit denen man sich bereits auf Formulierungen geeinigt hat, diese auch konsistent verwenden.

Im Rahmen des Interactive Alignment Ansatzes werden die facettenreichen Angleichungsprozesse auf linguistischer und non-linguistischer Ebene als **Instrumente erfolgreicher Kommunikation** beschrieben (Pickering & Garrod, 2004). Die linguistische Angleichung steht in direktem Zusammenhang mit dem Aufbau eines geteilten Wissensstandes und wird sowohl sprecher- als auch empfängerseitig als Akzeptanzevidenz gesehen. Denn wird die gleiche Formulierung erneut verwendet, so kann davon ausgegangen werden, dass der Empfänger die Intention des Sprechers verstanden hat und diese als Bestandteil des CG akzeptiert wird. Insbesondere die Vereinheitlichung der Wortwahl lässt nicht viel Spielraum für unterschiedliche Interpretationen. So konnte in Studien zu zwischenmenschlichen Dialogverhalten belegt werden, dass das lexikalische Alignment hoch mit dem Aufgabenerfolg korreliert (vgl. Nenkova, 2008; Reitter et al., 2007). Das Konzept des Alignments kann somit als ein weiteres Element des Groundings bezeichnet werden (Brennan, 1998).

An dieser Stelle soll in Kürze auf Prozesse eingegangen werden, die als **Ursachen** für die linguistische Angleichung diskutiert werden. Im Rahmen des Interactive Alignment Ansatzes wird davon ausgegangen, dass eine Angleichung der Gesprächspartner auf den verschiedenen Ebenen der linguistischen Repräsentationen unvermittelt, größtenteils automatisch und ohne bewusste Entscheidungen abläuft. Die Aktivierungsprozesse, die durch das Verarbeiten einer Äußerung automatisch bestehen, werden nicht durch Annahmen, die der Sprecher über seinen Gesprächspartner hat, beeinflusst, sondern gehen auf die **unvermittelten Mechanismen des Primings**[9] von Repräsentationen und Prozessen zurück.

In der psycholinguistischen Forschung ist es ein etablierter Befund, dass die sprachproduzierenden und -verarbeitenden Prozesse stark miteinander verbunden sind und durch zuvor verarbeitetes Sprachmaterial beeinflusst werden. Um beispielsweise das Wort *Sprachdialogsystem* zu interpretieren, muss der lexikalische Eintrag *Sprachdialogsystem* im mentalen Lexikon aktiviert werden (Meyer & Schvaneveldt, 1971). Eine solche Aktivierung eines Wortes oder auch einer

[9] Priming bezeichnet die Beeinflussung der Reizverarbeitung durch vorangegangene Wahrnehmungen.

grammatikalischen Struktur fällt nicht sofort wieder ab. Die bestehende Aktivierung erhöht die Wahrscheinlichkeit, dass der Rezipient anschließend die jeweilige linguistische Form verwendet (Pickering & Branigan, 1998).

Darüber hinaus wurde bereits 1967 mit der **motorischen Theorie der Sprachwahrnehmung** (Liberman et al.) eine Verbindung zwischen Sprachproduktion und -wahrnehmung postuliert. Die Grundannahme ist, dass das Hören eines bestimmten Sprachlautes einerseits die motorischen Mechanismen aktiviert, die für die Produktion des Lautes verantwortlich sind und andererseits Mechanismen, die die Wahrnehmung dieses Lautes ermöglichen. Der motorische Akt des Sprechens und der perzeptuelle Akt des Hörens sind so eng miteinander verbunden, dass es plausibel scheint, dass die dahinterliegenden Mechanismen ebenfalls eine enge Verbindung aufweisen (Goldstein, 1997). Die Entdeckung der **Spiegelneuronen**[10] bestätigt die vorhergesagte Verbindung zwischen motorischer Aktivität und Wahrnehmung. Insbesondere Rizzolatti & Craighero (2004) konnte deren Zusammenhang mit Sprache belegen. Die neurologische Befundlage stützt die Annahme, dass Alignment als ein fundamentaler Aspekt der Sprachverwendung bezeichnet werden kann.

Einige Untersuchungen können belegen, dass die ablaufenden Angleichungsprozesse den Gesprächsteilnehmern in den meisten Fällen nicht bewusst sind und nahezu ressourcenfrei ablaufen (Pickering & Garrod, 2004). Fast immer findet eine Synchronisation ohne Verhandlungsprozess statt (ebd.). Fowler et al. (2003) konnten nachweisen, dass eine Angleichung auf akustischer und phonetischer Ebene sehr schnell und beinah ohne mentale Beanspruchung geschieht. Es konnte weiterhin gezeigt werden, dass das Ausmaß der lexikalischen Angleichung viel stärker ist, wenn sich die Sprecher sofort auf den Begriff beziehen, als wenn sie dies nach einem gewissen zeitlichen Abstand tun (Brennan, 1996). Das lässt die Schlussfolgerung zu, dass die linguistische Konvergenz ein automatischer Prozess ist, der wahrscheinlich in großem Ausmaß durch Priming-Prozesse gesteuert wird. Durch das Alignment werden die mentalen Prozesse der Sprachproduktion und des Sprachverstehens der Gesprächspartner mit- und aneinander gekoppelt. Diese kognitiven joint actions vereinfachen die Prozesse der Produktion und des Verstehens in Dialogen stark (Pickering & Garrod, 2004). Eine Repräsentation, die gerade zum Verstehen einer Nachricht

[10] Spiegelneuronen sind Nervenzellen, die im Gehirn von Primaten während der Betrachtung eines Vorgangs die gleichen Potentiale evozieren, wie sie entstünden, wenn dieser Vorgang nicht bloß passiv betrachtet, sondern selbst durchgeführt werden würde. Eine von Mukamel et al. (2010) publizierte Studie berichtete über den ersten direkten Nachweis von Spiegelneuronen auch beim Menschen.

konstruiert wurde, kann sofort für die Produktion einer Antwort verwendet werden (oder umgekehrt). Somit kann nicht nur der Formulierungsaufwand einer Äußerung durch die drastische Reduzierungen möglicher Alternativen verringert werden, sondern auch der Verarbeitungsaufwand auf Seiten des Empfängers durch die identische Wortwahl. Im Bereich des Groundings können Kosten hinsichtlich der Produktion, der Rezeption und des Verständnisses gespart werden.

Es gibt allerdings auch Grund zur Annahme, dass es sich bei der Angleichung um einen **vermittelten Prozess** im Sinne des Rezipienten Designs handelt. Wie in Kapitel 3.1.2.3 beschrieben, kann beobachtet werden, dass Sprecher sich für Ausdrücke entscheiden, von denen sie ausgehen, dass sie die Wahrscheinlichkeit für eine erfolgreiche Kommunikation erhöhen. Kann ein Sprecher beispielsweise zwischen zwei Begriffen wählen, so wird er sich für den entscheiden, von dem er glaubt, dass sein Gesprächspartner ihn kennt und versteht. Die eigenen Erfahrungen mit diesem bestimmten Gesprächspartner stellen hierbei einen guten Anhaltspunkt dar, um dessen Wissensstand und linguistische Kompetenz einzuschätzen (Clark, 1996). So kann der Sprecher davon ausgehen, dass der Empfänger eine bestimmte Formulierung verstanden hat, wenn er diese zu einem früheren Zeitpunkt selbst verwendet hat. Als Ursache der Angleichung können demnach die geschätzten Kompetenzen seines Dialogpartners gelten.

Eine weitere mögliche Ursache des Alignments stellt eher einen **strategischen Ansatz** zur Verbesserung der sozialen Beziehung zwischen zwei Gesprächspartnern dar. So kann das bewusste Angleichen der Äußerungen dazu genutzt werden, um Gemeinsamkeiten zu verdeutlichen und die positive Einstellung des Gesprächspartners zu erhöhen. Unabhängig von der Richtung der Anpassungen konnte nachgewiesen werden, dass Alignment durch soziale Faktoren, wie Beziehungen oder Macht, beeinflusst wird (Giles et al., 1991). Viele Untersuchungen können belegen, dass das Spiegeln des Verhaltens des Gesprächspartners in Dialogen **affektive Konsequenzen** hat. Chartrand und Bargh (1999) demonstrierten, dass Konföderierte positiver beurteilt wurden, wenn sie die Körperhaltungen und Bewegungen ihrer Gesprächspartner kopierten. Berater, die die Haltung ihrer Klienten spiegelten, wurden sogar empathischer eingeschätzt (Maurer & Tindall, 1983). Ähnliche Effekte können auch im linguistischen Bereich nachgewiesen werden. So wurden Sprecher, die eine Angleichung des Wortschatzes zeigten, positiver beurteilt (Bradac et al.,1988). Van Baaren et al. (2003) konnte darüber hinaus zeigen, dass Kellnerinnen, die die Bestellung des Gastes exakt wiederholten, mehr Trinkgeld erhielten. Interessant ist auch der Befund, dass Leute, die in ihren Verhaltensweisen kopiert wurden, ein größeres prosoziales Verhalten zeigen (2004). Dagegen wird die Wahl von alternativen Äußerungen häufig als unhöflich oder korrigierend empfunden und kann im Empfänger negative Affekte auslösen (Branigan et al., 2010). Auf Basis dieser Befundlage wird die Konvergenz häufig auch als der "soziale Kitt" beschrieben, der hilft soziale Beziehungen zu stärken und strategisch angewendet werden kann.

Es kann davon ausgegangen werden, dass alle vorgestellten Ursachen parallele Gültigkeit besitzen. Alignment in einem bestimmten Kontext kann mehr oder weniger beide (vermittelte und unvermittelte) Prozesse enthalten. So kann davon ausgegangen werden, dass neben einem gewissen initialen Rezipientendesign zur Erhöhung der Kommunikationseffizienz, auch die unvermittelten Priming-Prozesse auftreten (Haywood et al., 2005). Je nach situativen Gegebenheiten wird sich die Gewichtung der Prozesse allerdings unterscheiden. Die vermittelten Komponenten gewinnen immer dann an Bedeutung, wenn in dem spezifischen Kontext eine effektive Kommunikation besonders wichtig oder gefährdet ist. Wobei diese Ursache des Alignments, die den kognitiv beanspruchenden Aufbau eines Empfängermodells beinhaltet, in durchschnittlichen zwischenmenschlichen Dialogen als weniger bedeutend erachtet wird. Da es zwischen zwei erwachsenen Personen aus demselben Kulturkreis normalerweise keine offensichtlichen Gefahren gibt, dass eine Kommunikation nicht erfolgreich verläuft, sind die unvermittelten Primingprozesse sehr viel wahrscheinlicher, nicht zuletzt, da diese ressourcenschonender sind (Horton & Keysar, 1996).

Nachdem nun drei Kernaspekte des Groundings eingeführt wurden, sollen abschließend Rahmenbedingungen einer Gesprächssituation beschrieben werden, die den Aufbau einer geteilten Wissensbasis mit minimalen Aufwandskosten erlauben. Sie sind insbesondere im späteren Verlauf der Arbeit von Interesse, wenn die Grounding-Prozesse mit einem SDS erleichtert werden sollen.

3.2.4 Optimale Bedingungen für Grounding-Prozesse

Der primäre und damit elementarste Gebrauch von Sprache findet sich in den Gesprächssituationen, in denen Sprecher und Empfänger einen Dialog von Angesicht zu Angesicht führen (Fillmore, 1981, S. 152; Clark, 1996). Die face-to-face Konversation zeigt nach Clark und Brennan (1991) die in Tabelle 4 dargestellten Charakteristiken.

Allerdings ist das Medium der Sprache in face-to-face Situationen vergänglich und es findet keine Aufzeichnung der Inhalte statt. Um das gegenseitige Verständnis dennoch sicherzustellen, werden diese Mängel kompensiert. So tritt in solchen Gesprächen beispielsweise eine erhöhte **Redundanz** der Inhalte auf (Salmen, 2002). Um Störgeräusche auszugleichen und den Empfang der Äußerung zu begünstigen, werden Inhalte wiederholt oder über mehrere Kanäle dargeboten. Die Sprecher nehmen hierbei einen höheren Produktionsaufwand in Kauf, um die Gegebenheiten der Gesprächssituation auszugleichen, die Rezeptionskosten für den Empfänger zu minimieren und die Verständniswahrscheinlichkeit zu erhöhen.

Tabelle 4: Eigenschaften der face-to-face-Kommunikation

Unmittelbarkeit	Kopräsenz	Zeitgleiche Anwesenheit beider Gesprächspartner.
	Sichtbarkeit	Die Gesprächspartner sehen sich.
	Hörbarkeit	Die Gesprächspartner hören sich.
	Direktheit	Die Gesprächspartner erleben die Aktion des Gegenübers ohne Verzögerung.
Medium	Vergänglichkeit	Das Medium ist vergänglich. Es verblasst schnell.
	Aufzeichnungs-freiheit	Die Aktionen der Gesprächspartner hinterlassen keine beständigen Aufzeichnungen.
	Simultanität	Ein zeitgleiches Produzieren und Verstehen ist möglich.
Kontrolle[11]	Vorbereitungs-freiheit	Die Gesprächspartner formulieren und artikulieren ihre Äußerungen unvorbereitet.
	Selbstbe-stimmtheit	Die Gesprächspartner bestimmen selbst, welche Aktionen wann ausgeführt werden.
	Selbstdarstel-lung	Die Gesprächspartner vollziehen Aktionen als sie selbst.

Andere Gesprächssituationen zeigen ebenso Mängel in der Unmittelbarkeit oder in der Vermittlung über das Medium (siehe Tabelle 5). Menschen verfügen über spezielle Techniken, um das gegenseitige Verständnis dennoch herstellen zu können. So werden Gesprächspartner am Telefon beispielweise direkt auf die Gesprächssituation hinweisen (*Ich fahre gerade!*), wenn diese den weiteren Verlauf der Kommunikation beeinflusst.

Tabelle 5: Konversationsmedien und ihre Gegebenheiten

Medium	**Rahmendbedingungen**
face-to-face	Kopräsenz, Sichtbarkeit, Hörbarkeit, Direktheit, Simultanität
Telefon	Hörbarkeit, Direktheit, Simultanität
Email	Aufzeichnung

[11] Die Kategorie der Kontrolle kann für die vorliegende Arbeit außer Acht gelassen werden, da ausschließlich von Interaktionen ausgegangen wird, die ein hohes Maß an Selbstbestimmtheit und -darstellung und dabei keine Vorbereitung voraussetzen.

Auch längere Pausen werden stets angekündigt, um Unsicherheiten beim Gesprächspartner zu vermeiden. Die mangelnde Sichtbarkeit wird durch spezifische Strategien (wie eine ausführlichere Situationsbeschreibung) ausgeglichen.

Diese Rahmenbedingungen der Gesprächssituation beeinflussen direkt die Aufwandskosten des Gesprächs. Je nach nutzbaren Wahrnehmungskanälen (Hörbarkeit, Sichtbarkeit), zeitlicher Kopräsenz und der Flexibilität der Dialogstruktur, entstehen unterschiedliche Kosten für die Beiträge der Gesprächspartner (Clark & Brennan, 1991). So sind beispielsweise die Kosten für einen Sprecherwechsel und Empfangsbestätigungen bei der Kommunikation via Email durch die mangelnde Unmittelbarkeit und Simultanität höher als bei einer face-to-face Kommunikation. Die Gesprächspartner werden versuchen, sich an diese Rahmenbedingungen anzupassen und möglichst eindeutige und umfangreiche Informationen in einer Email zu vermitteln, um häufige Sprecherwechsel und Verständnisprobleme zu vermeiden.

Eine Konversation, die **optimale Bedingungen** für eine erfolgreiche Etablierung eines geteilten Wissenstandes bei geringen Aufwandskosten bietet, erfüllt einerseits alle Faktoren der Kategorie Unmittelbarkeit (Clark & Brennan, 1991). Andererseits spielt neben der Kopräsenz, Sichtbarkeit, Hörbarkeit und Direktheit, auch die Kategorie des Mediums eine wichtige Rolle. So kann eine Inhaltsaufzeichnung die Vergänglichkeit des Mediums reduzieren und die Nachricht kann später noch angeschaut werden. Diese Ausweitung reduziert die Rezeptions- und Verständniskosten und kann im Falle eines Missverständnisses die Reparaturkosten senken. Die folgende Tabelle 6 soll die unterschiedlichen Gesprächskosten nach Clark und Brennan (1991) beschreiben und verdeutlichen, ob sie für den Empfänger, den Sprecher oder beide anfallen.

Tabelle 6: Aufwandskosten im Dialog

Konstruktions-/ Formulierungskosten	Sprecher
Exekutions-/ Produktionskosten	Sprecher
Rezeptionskosten	Empfänger
Verständniskosten	Empfänger
Gesprächsinitiierungskosten	Beide
Wartekosten	Beide
Asynchronitätskosten	Beide
Sprecherwechselkosten	Beide
Zeigekosten / Referenzkosten	Beide
Fehlerkosten	Beide
Reparaturkosten	Beide

3.3 Fazit

Um eine Orientierung bei der Gestaltung von Sprachdialogsystemen zu erlangen, wurde in den vorangegangenen Kapiteln der denkbar nutzerfreundlichste Dialog, genauer gesagt, der Dialog mit einem menschlichen Partner, näher beleuchtet. Dazu wurde die Collaborative Theory als theoretische Grundlage dieser Arbeit eingeführt und anhand der Konversationsmaximen nach Grice (1975) die prozeduralen Ziele einer Interaktion definiert. Den Phasen der Sprachproduktion wurden korrespondierende Phasen des empfängerseitigen Verständnisses gegenübergestellt und der kollaborative Charakter der Kommunikation verdeutlicht. Die joint actions des Sprechens und Empfangens können durch eine Zwei-Phasen-Struktur abgebildet werden. In der Akzeptanzphase signalisiert der Empfänger dem Sprecher, dass er die zuvor formulierte Äußerung der Präsentationsphase verstanden hat und fügt sie somit zu der geteilten Wissensbasis (CG) hinzu. Dieser Prozess zur Koordination von Wissensständen (Grounding) verdeutlicht, dass sowohl Sprecher als auch Empfänger zu gleichen Teilen verantwortlich für das erfolgreiche Abschließen einer Interaktion sind. Die von beiden Partnern geforderte Form, Stärke und Menge von Verständnisbekundungen im Rahmen der Akzeptanzphase ist je nach gegebenem Zweck der Kommunikation und Medium variabel. Das sogenannte Grounding Criterion definiert, welches Mindestmaß an Sicherheit gefordert wird, bevor der nächste Dialogschritt eingeleitet wird. Beide Kommunikationspartner tragen dazu bei, dass die anfallenden Aufwandskosten minimal sind bzw. im ausgewogenen Verhältnis zum Nutzen stehen (Least Collaborative Effort-Modell). Als weitere Kommunikationsstrategie, die Prozesskosten sparen kann und als Akzeptanzevidenz gilt, wurde die linguistische Angleichung der Gesprächspartner beschrieben (Alignment).

Als wesentlicher Erfolgsfaktor der zwischenmenschlichen Dialoge wurde somit der kollaborative Charakter der Konversation identifiziert und drei Aspekte des Groundings beschrieben, die den Aufbau der geteilten Wissensbasis unterstützen und damit ganz wesentlich zu gelungener Kommunikation beitragen. Im folgenden Kapitel soll gezeigt werden, dass die Collaborative Theory auch im Mensch-Maschine-Dialog Anwendung finden kann. In Bezug auf die Aspekte Feedback, Grounding Criterion und Alignment soll eine Statusanalyse aktueller SDS erfolgen aus der Optimierungspotentiale identifiziert werden können. Darauf aufbauend wird das Potential und die Umsetzung der Übertragung jener Grounding-Elemente diskutiert.

4 Kollaborative Mensch-Maschine-Dialoge

Nachdem in Kapitel 3 wesentliche Prinzipien der menschlichen Dialogführung dargestellt wurden, sollen nun aktuelle SDS im Hinblick auf die eingeführten Grounding-Elemente untersucht werden. In Anlehnung an Clarks Thesen soll die Interaktion mit einem SDS dabei als eine kollaborative Aktion verstanden werden, bei der ein Gesprächsteilnehmer durch eine Maschine repräsentiert wird. Die Prozesse des Sprechens und Verstehens werden dabei, in Anlehnung an den zwischenmenschlichen Dialog, als partizipatorische Handlungen gesehen, die zusammen die *joint action* der Kommunikation bilden.

Im ersten Abschnitt des Kapitels wird veranschaulicht, dass Menschen Strategien der Kollaboration auch im Rahmen einer Konversation mit einem maschinellen Gesprächspartner anwenden. Danach wird im Rahmen einer Statusanalyse aufgezeigt, dass bei der bisherigen Gestaltung von SDS der kollaborative Aspekt der Kommunikation nahezu außer Acht gelassen wurde. Der hohen Responsivität und Anpassungsleistung der Nutzer stehen mangelnde Grounding-Prozesse des SDS gegenüber. Die Folgen eines Dialogdesigns, welches die nutzerseitigen Erwartungshaltungen bezüglich Feedback, Alignment und Flexibilität des Grounding Criterions ignoriert, werden verdeutlicht.

Folglich wird in Kapitel 4.2 argumentiert, dass eine solche Übertragung der Konversationsmaximen und kollaborativen Annahmen auf den Mensch-Maschine-Dialog zulässig und vielversprechend sind. Auf Basis der Forschungslage endet das Kapitel mit einer deutlichen Empfehlung, Maschinen als kollaborative Gesprächspartner zu konzipieren, um Dialoge mit SDS weniger belastend und frustrierend zu gestalten.

4.1 Grounding-Prozesse in Mensch-Maschine-Dialogen

Indizien für die Verwendung kollaborativer Strategien zeigen sich neben Mensch-Mensch-Interaktionen auch bei den Mensch-Maschine-Interaktionen. So nehmen Nutzer häufig eine erhöhte Anpassungsleistung in Kauf, um das Erreichen des Kommunikationsziels nicht zu gefährden und die Effizienz zu steigern. Inwiefern Menschen ihr Verhalten an ihre eigenen Vermutungen über die Systemkompetenz anpassen, soll in den folgenden Absätzen erläutert werden.

Einerseits zeigen Menschen beim Sprechen mit Maschinen deutliche Abweichungen von der zwischenmenschlichen Standardsprache, um die Effektivität der Interaktion zu gewährleisten (z. B. Baber & Strammers, 1989; Fischer, 2006; Fraser & Gilbert, 1991). Bereits Zoeppritz (1985) prägte für die bei Mensch-Maschine-Dialogen systematisch auftretenden Sprachanpassungen den Begriff **Computer Talk** und fügte damit ein weiteres Sprachregister (Kap. 3.1.2.3) ein. Computer Talk definiert sich nach Krause (1992):

„[…] wie sich Menschen ausdrücken und verhalten, wenn sie Computersystemen, die natürlich-sprachliche Eingaben zulassen, Anweisungen geben." (ebd., S. 1)

Vereinfachend können die Abweichungen von der Standardsprache in die zwei Kategorien **Simplification** (dt. Vereinfachung) und **Clarification** (dt. Verdeutlichung) eingeteilt werden. Im Sinne der Vereinfachung verletzen Nutzer Grammatikregeln und missachten ihr eigenes lexikalisches Wissen. So zeichnen sich Spracheingaben häufig durch fehlende Artikel, fehlende Verben und Überleitungen aus. Der Aspekt der Verdeutlichung führt zu einem reduzierten Anteil an impliziten Informationen. Um die Verständniswahrscheinlichkeit zu erhöhen, werden vermehrt explizite Formulierungen gewählt (Krause & Hitzenberger, 1992; Richards & Underwood, 1984). Auch Boyce (2008) konnte einen erhöhten Gebrauch von Kommandosprache[12] und eine geringe Wortanzahl im Dialog mit Maschinen feststellen.

Diese nutzerseitigen Anpassungen können einerseits damit erklärt werden, dass Nutzer von einer geringen sprachlichen Kompetenz des Systems ausgehen (Krause, 1992). Um diese Schwächen des Gesprächspartners zu kompensieren und das Interaktionsziel zu erreichen, finden diese Vereinfachungs- bzw. Verdeutlichungsstrategien Anwendung.

Andererseits kann die bewusste Steigerung der **Effizienz** durch knappe Spracheingaben auch eine Begründung der Vereinfachungs- und Verdeutlichungsstrategie sein. Graham et al. (1999) konnte im Fahrtkontext beobachten, dass insbesondere Nutzer ohne Erfahrung (Novizen) zur sprachlichen Bedienung verschiedener Infotainment- und Navigationsfunktionen in über 90% der Fälle Kommandos verwendeten. Über die Hälfte dieser Befehle bestand aus ein oder zwei Worten. Die Probanden begründeten die Verwendung knapper Spracheingaben allerdings nicht mit der geringen antizipierten Sprachkompetenz des Systems, sondern damit, dass sie die Aufmerksamkeit und Zeit für die Sprachbedienung reduzieren wollten, um sich besser auf die Fahraufgabe konzentrieren zu können (vgl. ebd., S. 317-320). Die reduzierte Kommunikation über Kommandos begründeten sie demnach durch den Wunsch nach einem raschen Erreichen des Bedienziels. Engelhardt (2009) stützt diese Argumentation und konnte im Rahmen einer Fahrsimulatorstudie nachweisen, dass unabhängig von der angenommenen Sprachkompetenz, die

[12] Kommandosprache ist eine Sprechweise, die durch definierte Wörter oder Sätze eine eindeutige Handlungsaufforderung zum Ausdruck bringt.

Spracheingabe zugunsten einer raschen Erreichung des Bedienziels auf Kommandos reduziert wird.

Die Befundlage lässt vermuten, dass die **Ursache des Computer Talks** nicht allein in dem expliziten Einbezug des CG in Planungsphase liegt. Insbesondere im Rahmen der mental beanspruchenden Fahrsituation machen sich Nutzer nicht die Mühe, die geschätzte Sprachkompetenz des Empfängers in ihre Formulierung einzubeziehen. Vielmehr verkürzen sie ihre eigene Spracheingabe aus dem egozentrischen Gesichtspunkt heraus, weniger Zeit und damit Ressourcen in die eigene Sprachkonstruktion zu verwenden und vollziehen den Adjustierungsprozess nur wenn nötig.

In Situationen, in denen ausreichend mentale Ressourcen zur Verfügung stehen, kann allerdings beobachtet werden, dass neben dem Monitoring and Adjustment Model auch eine explizite Bezugnahme auf den CG geschieht. So lässt sich im spielerischen Umgang mit Avataren, die dazu in der Lage sind, gesprochene Sprache zu verstehen, häufig ein Austesten der Systemgrenzen durch den Nutzer beobachten. So wurde der Museumsavatar Max in einer Studie von Kopp (2008) neben angebrachten, regulären Anfragen auch mit Testfragen (19.6%), Fragen zu dem System (14.6%) und Beschimpfungen (11.1%) konfrontiert. Die Nutzer zeigten ein implizites Testen der Intelligenz und Kompetenz des Avatars, in dem sie ihn mit einer Fremdsprache ansprachen, ihm Fragen auf einer Meta-Ebene stellten (*Wer hat dich gebaut?*) oder ihn mit menschlichen Anfragen konfrontierten (*Bist du verliebt?*). Dieses Explorieren entspricht zwar nicht dem Verhalten eines kooperativen Gesprächspartners nach Grices Maximen, aber dafür umso mehr der expliziten Etablierung eines Common Grounds aus Gründen der Neugier. Man kann also beobachten, dass es in Kontexten, in denen mentale Kapazitäten ausreichend verfügbar sind und es nicht um eine gemeinsame Zielerreichung geht, die nutzerseitige Bereitschaft zur Kollaboration größer ist.

Viele Anwendungsbeispiele können verdeutlichen, dass Nutzer einen **Common Ground** in Interaktionen mit Maschinen voraussetzen und verwirrt reagieren, wenn sie auf jenen nicht referenzieren können. Wie in zwischenmenschlichen Dialogen gehen sie davon aus, dass Inhalte, die bereits bestätigt wurden, als fester Bestandteil des CG gelten und nicht erneut eingeführt werden müssen. Wie stark diese logischen Ableitungen der geteilten Wissensbasis in der Konversation mit Maschinen verankert ist, lässt sich u.a. an der Verwendung von Ellipsen zeigen. Ohne sich der technologischen Herausforderung bewusst zu sein, werden auch in Dialogen mit Maschinen Satzteile ausgelassen, die zuvor Erwähnung fanden (vgl. Gieselmann & Stenneke, 2006).

Ein Beispiel dafür, dass der sprachliche oder situative Kontext vorausgesetzt wird, ist folgender Dialog (insbesondere der letzte Turn):

Nutzer	Helga Otto mobil anrufen.
System	Möchten Sie Helga Otto mobil anrufen?
Nutzer	Ach so, nee, privat.

Auch das Senden und Empfangen von Statusindikatoren oder **Feedback** sind Vorgänge, die den Dialog zwischen Menschen kontinuierlich begleiten und zu großen Teilen für das erfolgreiche Grounding und das Erreichen des Kommunikationsziels verantwortlich sind (siehe Kap. 3.2.1). Die gesamte Interaktion profitiert von den Rückmeldungen des Empfängers, die dem Sprecher helfen zu diagnostizieren, in welchem Zustand sich der Gesprächspartner befindet. Dies ermöglicht wiederrum die Äußerungen an den Empfänger kontinuierlich anzupassen. Da die Rückmeldungen somit ein essentieller Bestandteil der zwischenmenschlichen Kommunikation sind, überrascht es nicht, dass sowohl das Senden, als auch das aktive Empfangen von Feedback, auch im Rahmen von Mensch-Maschine-Interaktionen, von Nutzern beobachtet werden kann. In dem folgenden Kapitel sollen exemplarisch Beobachtungen aufgeführt werden, die verdeutlichen, dass Menschen versuchen, dieses Element des Groundings auch in Dialogen mit Maschinen anzuwenden. Weiterhin werden gegenwärtige automotive SDS im Hinblick auf die Bandbreite ihrer Rückmeldungen analysiert werden.

4.1.1 Feedback automotiver SDS

Obwohl Menschen auch dann eine hohe Responsivität zeigen, wenn sie zu sprachverarbeitenden Maschinen sprechen, unterstützt das Design aktueller Sprachdialogsysteme nur selten die nutzerseitige Äußerung von Empfangsbestätigungen (Ward & Heeman, 2000). Sollten Nutzer prozessbegleitende **Backchannels** produzieren, so werden diese in den meisten Fällen ignoriert oder führen gar zu einer Verlängerung des Dialogs. Denn die wenigen Systeme, die einen Barge-In erlauben, reagieren auf eine Nutzereingabe während einer Systemausgabe mit einem sofortigen Stopp des Prompts. Noch bevor die Nutzereingabe interpretiert wurde und festgestellt werden konnte, ob es sich dabei nur um eine prozessbegleitende Empfangsbestätigung handelt, wird die Äußerung des Systems angehalten und ein Sprecherwechsel initiiert, der so vom Nutzer nicht vorgesehen war. Bei dieser Systemkonfiguration stört das Produzieren von Backchannels sogar den Mensch-Maschine-Dialog, obwohl es den zwischenmenschlichen Grounding-Prozess begleitet und unterstützt.

Bezüglich der **dialogstrukturierenden Gliederungssignale**, die das Ende oder den Beginn einer Äußerung markieren, berichtet Porzel (2006), dass diese Signale sprachverarbeitenden Ma-

schinen auch dann entgegengebracht werden, wenn diese darauf keine Reaktion zeigen. Bisher werden die Zeitpunkte zum Sprecherwechsel in den meisten SDS durch die technologischen Restriktionen des Systems vorgegeben. Eine kooperative Übergabe oder Übernahme des Turns durch den Nutzer ist nicht vorgesehen. Die nutzerseitige Produktion von dialogstrukturierenden Signalen ist somit (aus heutiger Systemsicht) überflüssig.

Dennoch können Studienergebnisse belegen, dass ein Bedarf der Nutzer besteht, diese Acknowledgments auch im Mensch-Maschine-Dialog zu zeigen. Ward und Heeman (2000) implementierten und evaluierten ein SDS, dass den Nutzern die Möglichkeit bot Empfangsbestätigungen zu äußern. Auch ohne eine spezifische Instruktion konnten sie beobachten, dass über die Hälfte der Probanden diese Form der Empfangsbestätigungen mindestens einmal und 29% der Probanden sie umfassend über alle Dialoge hinweg nutzen. Dieses knappe Drittel der Probanden bevorzugte insbesondere Acknowledgments vor einer weiteren Kommandoeingabe. Diese Befunde werden von Okato et al. (1998) gestützt und um die Erkenntnis erweitert, dass Menschen vor allem dann mehr Backchannels produzieren, wenn das System dieses Verhalten auch zeigt.

Grundsätzlich kann geschlossen werden, dass Menschen auch im Dialog mit Maschinen ihrer Verantwortung als Empfänger nachkommen und Feedbacksignale produzieren, auch wenn diese ignoriert werden. Im Sinne der Kollaboration erwarten Menschen diese Rückmeldungen auch von ihrem Gesprächspartner und verlangen sowohl während als auch nach ihrer Spracheingabe Bestätigungen darüber, dass sie verstanden wurden (Brennan, 1998).

So konnten Wrede et al. (2006) bei der Analyse von Kommunikationen mit einem Roboter zeigen, dass Menschen den elektronischen Gesprächspartner besser bewerteten, wenn dieser eine hohe Initiative zeigte und aktiv sehr viel Feedback gab. Mangelte es den Nutzern allerdings an Feedback, so wiederholten sie nach einer kurzen Wartephase die Spracheingabe, modifizierten die Akzentuierung, begannen zu buchstabieren, verwendeten Synonyme, zeigten mimische Variationen oder schnippten mit dem Finger, um so den Zugang zu dem Gesprächspartner zu erneuern und seine Aufmerksamkeit zurück zu erlangen. Die Nutzer forderten damit aktiv Feedback, wenn der Roboter dieses selbstständig nicht zeigte. Auch Gieselmann und Stenneke (2006) konnten als eine häufige Form der Metakommunikation Klärungsfragen an den Versuchsleiter beobachten (z. B. Hat der Roboter mich verstanden?), die auch darauf hinweisen, dass es den Nutzern vor allem an Verständnisevidenzen fehlte.

Im Vergleich zu der hohen Initiative und Responsivität, die Menschen in Kommunikationen zeigen, bieten auch aktuelle Sprachdialogsysteme im Fahrzeug nur **unzureichend Feedback** über die ablaufenden Empfangsprozesse. Üblicherweise zeigen sie keinerlei Reaktionen bis der Nutzer seine Eingabe beendet hat. Da somit keine Verständnisevidenzen korrespondierend zu den sprachproduzierenden Handlungen des Nutzers gezeigt werden, kann nur anhand der Sys-

temantwort antizipiert werden, welchen Dialogzustand das System erreicht hat. Die kooperative Verantwortung des Empfängers, dem Sprecher Indizien über die eigenen Verstehensprozesse zu präsentieren, erfüllen aktuelle SDS nur mangelhaft und erschweren somit die Koordination der Wissensstände und die Etablierung des CG.

"Today's typical spoken dialog system produces no response until after the speaker finishes an utterance. Humans, in contrast, are very responsive, reacting frequently while the speaker is talking." (Ward, 1996, S. 1728)

Die grafischen Bedienschnittstellen (z. B. Touchscreens; Abbildung 12) sind den Sprachdialogsystemen an dieser Stelle überlegen. Wie bereits in Kapitel 2.3.2 angedeutet, bieten verschiedene gegenwärtige grafische Bedienschnittstellen im automotiven Kontext eine exzellente Basis für Grounding-Prozesse, indem sie Resultate der Nutzereingabe kontinuierlich anzeigen und somit in den meisten Fällen eine externe Repräsentation des geteilten Kontextes darstellen können. So vermittelt beispielsweise das aktuelle grafisch-haptische Bedienkonzept von Volkswagen (RNS510) neben taktilen (gefühlter Tastendruck eines Hardkeys) und visuellen (Verdunkeln des virtuellen Touchbuttons bei Tastendruck), zusätzlich auditives Feedback (Klickgeräusch bei Tastendruck) während der Bedienung. Vergleicht man SDS zu dieser Schnittstelle, so mangelt es ihnen an den ad-hoc Repräsentationen von Handlungsfolgen (Brennan, 1998).

Abbildung 12: Beispiele für grafisch-haptische Bedienschnittstellen

Um SDS mit einem vergleichbaren Statusfeedback auszustatten adaptierten Brennan und Hulteen (1995, S. 144) das Contribution Modell des Groundings (vgl. Kap. 3.2) im Hinblick auf den Ablauf der maschinellen Spracherkennung. Sie erweiterten das Modell der Dialogzustände um vier weitere Stufen und identifizierten demnach acht mögliche Dialogzustände, die dem Nutzer anzeigen konnten, in welchem Zustand sich das SDS befindet und welche Handlungsoptionen ihm zur Verfügung stehen (siehe Tabelle 7).

Anhand dieses Modells soll im Folgenden eine Analyse und Bewertung einiger aktueller automotiver Sprachdialogsysteme erfolgen. Global betrachtet bieten all diese Bedienschnittstellen dem Nutzer nur eine stark reduzierte Form der Rückmeldung. In den meisten Fällen zeigen sie ledig-

lich nur den Zustandswechsel zwischen dem ausgeschaltetem System (Zustand 0) und der Erkenneröffnung (Zustand 1) durch eine akustische Tonfolge an.

Tabelle 7: Adaptiertes Contribution Modell

Zustand 0	**Das System ist nicht bereit.** Das System kann keine Eingabe vom Nutzer empfangen.
Zustand 1	**Das System ist bereit.** Das System kann eine Eingabe vom Nutzer empfangen.
Zustand 2	**Das System empfängt** eine Nutzereingabe.
Zustand 3	**Das System erkennt** eine Nutzereingabe (aber interpretiert sie noch nicht).
Zustand 4	**Das System interpretiert** die Eingabe.
Zustand 5	**Das System hat eine Handlungsabsicht** und hat die Eingabe als ein Kommando an den Dialogmanager übertragen.
Zustand 6	**Das System führt** das Kommando **aus**.
Zustand 7	**Das System berichtet**, ob es das Kommando erfolgreich ausführen konnte.

Im Infotainment-Bediensystems von Audi, dem **Audi MultiMedia Interface** (Audi MMI 3G) wird diese Tonfolge durch eine visuelle Unterscheidung zwischen den Aktivitätszuständen über eine Statusleiste unterstützt. Abbildung 13 verdeutlicht, dass lediglich ein Icon über den Aktivitätsgrad (Zustand 0 oder 1) des Systems informiert. Fährt man im Dialog fort, so erfolgt eine sprachliche Ankündigung der Handlungsabsicht (Zustand 5; z. B. „Route wird berechnet“).

Abbildung 13: Audi MMI 3G Aktivitätsanzeige

Erweitert wird die grafische Anzeige der Statusleiste zusätzlich durch eine Bedienhilfe, die kontextsensitiv sprechbare Kommandos anzeigt (engl. *speak what you see*). Dieser speak what you see Ansatz hat sich herstellerübergreifend durchgesetzt.

Auch das Infotainment-Bediensystems von BMW, das **BMW 750LI iDrive**, signalisiert den Übergang von Zustand 0 zu Zustand 1 (und umgekehrt) mit einem Ton. Unterstützend erscheint im

Kombidisplay ein Icon mit einer textuellen Aufforderung zur (erneuten) Spracheingabe (siehe Abbildung 14).

Abbildung 14: BMW 750LI iDrive System

Ähnliche Feedbackstrukturen lassen sich bei dem SDS von dem **Mercedes Benz S550** beobachten. Das System produziert einen Ton, nachdem die PTT Taste gedrückt wurde und das System seinen Aktivitätszustand ändert. Wenn das System aktiviert wurde, erscheint zusätzlich ein Icon in der Head-Unit (siehe Abbildung 15).

Abbildung 15: Mercedes Benz S550 SDS

Eine Besonderheit zeigt das Infotainmentsystems von Ford, das **Ford Edge SYNC System**, dass durch unterschiedliche Klänge signalisiert, ob ein Kommando verstanden wurde oder nicht. Es ist damit eines der wenigen Systeme, das dem Nutzer Verständnisevidenzen liefert und Zustand 3 kommuniziert. Eine ähnliche Form der Verstehensevidenz zeigt das **Honda Acura RL handsfreelink System**, das ein verstandenes Kommando mit einem Ton bestätigt, während es beispielsweise in ein Untermenü wechselt. Damit wird zusätzlich eine Handlungsausführung kommuniziert (Zustand 6).

Das SDS des **Lexus RX 450h** liefert die ausführlichste Rückmeldung des Zustands 3. Neben einer speak what you see Anzeige, vermittelt es über ein Icon den Aktivitätsstand des Systems und zeigt das vom Spracherkenner erkannte Ergebnis in einer Sprechblase an (siehe Abbildung 16).

Abbildung 16: Grafische Anzeige SDS Lexus RX 450h

Damit verwenden aktuelle Sprachdialogsysteme vorrangig Töne als dialogstrukturierende Signale. Je nach System fungieren sie als dialogstrukturierendes Signal oder als Empfangsbestätigung (Ford und Acura). Bei den meisten Systemen erfolgen die Zustände 6 und 7 unkommentiert durch eine grafische Anzeige des Kontextwechsels.

Tabelle 8 verdeutlicht zusammenfassend, welche Zustände des adaptierten Contribution Modells von derzeitigen SDS adressiert werden.

Tabelle 8: Statusanalyse Feedback SDS

	Audi MMI 3G	BMW 750LI iDrive	Mercedes Benz S550	Ford Edge SYNC	Honda Acura RL hands-freelink® System	Lexus RX 450h
Zustand 0: Das System ist nicht bereit.						
Zustand 1: Das System ist bereit.	X	X	X	X	X	X
Zustand 2: Das System empfängt.						
Zustand 3: Das System erkennt.				X	X	X
Zustand 4: Das System interpretiert.						
Zustand 5: Das System hat eine Handlungsabsicht.	X	X	X	X	X	X
Zustand 6: Das System führt aus.					X	
Zustand 7: Das System berichtet das Ergebnis.						

Einen weiteren kritischen Aspekt stellt der **Verarbeitungszustand** des Systems dar, der den Übergang von Zustand 2 zu Zustand 5 begleitet und bei umfangreichen Navigationszieleingaben mehrere Sekunden in Anspruch nehmen kann. Viele derzeitige Systeme bieten ausschließlich statische Informationen und kündigen die Verarbeitungsphase mit einem Prompt wie *Einen Moment bitte.* an. Leider enthält dies keine dynamische Information über die Dauer der Wartezeit. Auch das an dieser Stelle fehlende Feedback wird als Rückmeldung interpretiert. Ohne die transparente Vermittlung der Dialogzustände lässt die Stille des Systems zwei Interpretationen durch den Nutzer zu: entweder das System wartet auf einen Nutzerinput oder es verarbeitet die letzte Eingabe noch (Cathcart, 2003). Beide Zustände müssen für den Nutzer transparent dargestellt werden, da sie unterschiedliche Handlungsoptionen beinhalten. Während die Stille im ersten Fall als Aufforderung zur Spracheingabe verstanden werden soll, würde eine weitere Eingabe durch den Nutzer im zweiten Fall das Dialogsystem weiter belasten und eine Fehlerkennung wahrscheinlich machen.

Die Analyse konnte aufzeigen, dass SDS nur eingeschränkt Rückmeldungen darüber bieten, in welchem Dialogstatus sie sich gerade befinden. Ohne die umfassende Vermittlung des Systemzustandes sind Nutzern die verfügbaren Handlungsmöglichkeiten häufig nicht bekannt. Laut Brennan (2009) lassen sich viele Fehler, die in Mensch-Maschine-Dialogen auftreten, durch jenes fehlende Grounding-Element erklären. Der anschließende Abschnitt soll anhand empirischer Untersuchungen die Folgen der mangelnden Rückmeldungen von SDS verdeutlichen.

Folgen mangelnden Feedbacks

Verschiedene Studien können belegen, dass diese nutzerseitige Unsicherheit sich negativ auf die Bedienleistung auswirken kann.

So wurde eine **Vorstudie**[13] durchgeführt, um zu untersuchen, wie häufig Fehlbedienungen (TTE-Fehler) mit einem Seriensystem der Volkswagen AG auftreten. Während einer Realfahrtstudie auf dem betriebsinternen Prüfgelände sollte ermittelt werden, wie häufig Spracheingaben auftreten, noch bevor das System empfangsbereit ist. Versuchsteilnehmer waren 30 Volkswagenmitarbeiter, die nicht an der Entwicklung des Sprachdialogsystems beteiligt waren oder Erfahrung in der Bedienung automotiver SDS hatten. Die Altersspanne betrug 23 bis 58 Jahre (*MW*= 39.7 Jahre; *SD*= 8.3 Jahre). 50% der Probanden waren männlich.

[13] Unveröffentlichte, interne Studie der Volkswagen AG.

Das verwendete System war vergleichbar zu den Systemen der Statusanalyse (Kap. 4.1.1). Die Probanden führten 10 Sprachkommandos aus (z. B. „Käthe Sommer mobil anrufen.“). Auf den Druck der PTT-Taste folgte eine kurze Wartephase bis ein Jingle ertönte, die Anzeige S*prachbedienung aktiv* im Kombidisplay erschien und der Erkenner aufnahmebereit war. Während der sprachlichen Interaktion mit dem SDS waren die Versuchspersonen (VP) dazu angehalten einen konstanten Abstand (50 m) zu einem Vorderfahrzeug einzuhalten, welches zwischen 80 und 120 km/h fuhr.

Insgesamt sprachen 79% der Probanden das erste Kommando zu früh ein. Danach zeichnet sich eine deutliche Lernkurve ab. Nach durchschnittlich 3.2 Versuchen waren die Probanden in der Lage, den Beginn ihrer Spracheingabe an den Jingle anzupassen. Insgesamt konnte die Studie belegen, dass das aktuelle Seriensystem nur unzureichende Rückmeldung bietet, wann das System empfangsbereit ist. Nur 21% der Probanden verstanden das System ohne jede Instruktion. Versuchsleiterbeobachtungen konnten belegen, dass Nutzer insbesondere in dem Kombidisplay nach Feedback über den Dialogzustand des Systems suchten. Die zugrundeliegende Semantik des akustischen Signals, welches über den Dialogzustand informiert, musste erst in drei Interaktionen erlernt werden. Wie bereits in Kap. 2.3.3 erwähnt, konnte in Langzeitstudien beobachtet werden, dass Nutzer bereits nach zwei fehlerbehafteten Interaktionen die Modalität wechseln. In der Studie konnte darüber hinaus eine hohe Anpassung der Zweitaufgabenausführung an die aktuelle Fahrsituation beobachtet werden. Bei komplexeren Fahrmanövern, wie beispielsweise dem Aufschließen zu dem Vorderfahrzeug, wurde die Sprachbedienung nicht gestartet. Erst bei antizipierter geringer Komplexität der Primäraufgabe, starteten die Versuchspersonen den Sprachdialog.

Im Rahmen einer internen **Benchmark-Studie**[14] mit den automotiven Sprachdialogsystemen der Statusanalyse (Kap. 4.1.1) konnten ähnliche Effekte beobachtet werden. Über alle Systeme hinweg gemittelt sprachen 27.6% der 40 Probanden während das System nicht aktiv war oder selbst sprach. Obwohl alle vorgestellten Systeme ihr Aktivität (Status 0 oder 1) zurückmeldeten, sprachen 22.5% der Probanden ihr Kommando zu früh ein und produzieren damit TTE-Fehler.

Die hohen Gesprächsinitiierungskosten und die erhöhte Wahrscheinlichkeit zur Fehlbedienung durch das mangelnde oder unzureichende Feedback stellen damit eine ernstzunehmende Einschränkung der Schnittstelle dar. So erfordert es von den Probanden beider Studien eine gewisse

[14] Unveröffentlichte, interne Studie der Volkswagen AG.

mentale Beanspruchung, einen Sprachbeitrag zurückzuhalten und ihn nicht sofort nach Tastendruck einzusprechen.

Darüber hinaus kann beobachtet werden, dass Nutzer jede längere Pause (> 1 Sekunde), die durch das Sprachdialogsystem produziert wird, als eine Aufforderung zum Sprecherwechsel verstehen (Harris, 2005). Wird der Verarbeitungszustand nicht transparent vermittelt, so wird der Nutzer das Sprechen beginnen und damit einen Fehler des Dialogsystems begünstigen.

In einer Benchmark-Studie mit portablen, sprachbedienbaren Navigationsgeräten berichtete auch Goulati (2011) davon, dass Probanden bei fehlendem Feedback stark verunsichert reagierten. Diese mangelnde Transparenz oder das unerwartete Systemverhalten können negative Effekte, wie steigende Beanspruchung und Ablenkung zur Folge haben (Burnett & Joyner, 1997). So belegte Kainer (2007), dass das Fehlen von Feedback beim Nutzer eine mentale Überbelastung verursachte.

Somit kann geschlossen werden, dass aktuelle Sprachdialogsysteme ihrer Verantwortung als Empfänger nur begrenzt gerecht werden. Einerseits reagieren sie auf die vom Nutzer produzierten Empfangsbestätigungen nur unzureichend, andererseits zeigen sie selbst kaum Verständnisevidenzen. Im folgenden Kapitel soll verdeutlicht werden, dass auch die Aufgabe, den Gesprächsaufwand minimal zu halten, bei den aktuellen Konfigurationen als nicht zufriedenstellend gelöst bezeichnet werden kann.

4.1.2 Grounding Criterion automotiver SDS

Gegenwärtige automotive Sprachdialogsysteme liefern Akzeptanzevidenzen in den meisten Fällen explizit. In Form von inkrementellen Groundingstrategien lassen sie sich jede Nutzereingabe erneut bestätigen (Brennan & Hulteen, 1995) und zeigen damit unabhängig von dem Kommunikationszweck und der Gesprächssituation ein stark konservatives GC (vgl. Kap. 3.2.2). In dem folgenden Beispieldialog ist solch ein typischer Interaktionsablauf zur Zieleingabe dargestellt. Auch wenn sich in der n-best list der Erkennerergebnisse nur ein Eintrag befindet, muss der Nutzer jede inhaltstragende Eingabe explizit akzeptieren.

Nutzer	Adresse eingeben.	
System	Möchten Sie eine Stadt oder Straße eingeben?	Implizite Evidenz
Nutzer	Stadt eingeben	
System	Bitte nennen Sie die Stadt!	Implizite Evidenz
Nutzer	Braunschweig	

System	Meinten Sie Braunschweig?	Explizite Evidenz
Nutzer	Ja.	
System	Braunschweig wurde übernommen. Möchten Sie eine Straße eingeben oder die Zielführung starten?	
Nutzer	Straße eingeben	
System	Die Straße, bitte.	Implizite Evidenz
Nutzer	Karlstraße.	
System	Meinten Sie Karlstraße?	Explizite Evidenz
Nutzer	Ja.	
System	Karlstraße wurde übernommen. Möchten Sie eine Hausnummer eingeben oder die Zielführung starten?	
Nutzer	Zielführung starten.	
System	Die Routen werden berechnet. <Endjingle>	

Diese **defensive Interpretation** der Erkennerergebnisse (Kapitel 2.2.4) vermeidet potentielle Missverständnisse, zeichnet sich aber durch eine geringe Dialogeffizienz aus. Bezogen auf das Contribution Modell (vgl. Kap. 3.2) bedeutet dies, dass die Akzeptanzphase der Konversation durch die Bestätigungsanfragen, insbesondere bei korrekter Erkennung, unnötig verlängert wird. Aktuelle Seriensysteme erhöhen damit den kollaborativen Aufwand in unangemessener Höhe, widersprechen dem Ansatz des Least Collaborative Efforts und verletzen die Grice‘schen Maximen der Quantität und der Art und Weise (Grice, 1975). Dadurch stehen sie im starken Gegensatz zu zwischenmenschlichen Kommunikationsstrategien, die durch eine hohe Situationsabhängigkeit und Flexibilität der Interaktionsabläufe den gemeinsamen Aufwand reduzieren.

Neben der Vermeidung von Fehlerkennungen gehen die Bestätigungsanfragen auch auf das Gestalten von SDS in Anlehnung an grafische Nutzerschnittstellen und deren Konsistenzmaxim zurück (Nass & Brave, 2005). So werden auch sprachliche Interfaces immer so gestaltet, dass sie identische Dialogabläufe zeigen. In vielen Umsetzungen wird daher die Systemkonsistenz über die Dialogeffizienz gestellt.

Zugunsten der Effektivität und Konsistenz müssen Nutzer aktueller SDS somit eine geringe Effizienz im Dialogablauf in Kauf nehmen. Die Folgen dieser statischen Konfiguration sind dem folgenden Abschnitt zu entnehmen.

Folgen des statischen GC

Durch ein inflexibles konservatives GC werden die Dialogabläufe mit SDS von Nutzern in vielen Untersuchungen als unangemessen und **zu lang** eingestuft (Hipp et al., 2008). Dieser Befund wird auch durch eine umfangreiche Nutzerbefragung bestätigt (Vital, 2009). 365 Nutzer eines automotiven Sprachdialogsystems (Ford Sync) wurden befragt, welche Systemverbesserungen sie priorisieren würden. Dabei wurde insbesondere die Kategorie *kürzere Dialoge mit weniger Bestätigungsaufforderungen* als sehr wichtig bewertet. Befragt nach den Schwachstellen des aktuellen Systems wurden in Übereinstimmung dazu *zu lange Dialoge* und *zu viele Bestätigungen* genannt.

Das dauerhaft konservative GC führt nutzerseitig nicht nur zu einem geringen Effizienzempfinden, sondern auch zu einem **geringen Zuverlässigkeitsempfinden**. Denn häufig werden die redundanten Dialogschritte als Verständnisproblem des Systems fehlinterpretiert. Ein erneutes Explizieren der Informationen und damit des gesamten CG geschieht in zwischenmenschlichen Dialogen nur bei einem unzureichenden Verständnis. Somit werden die ständigen Nachfragen des Systems als Indikator für einen unzureichenden Dialogzustand verstanden.

Neben der Erhöhung der Wartekosten seitens des Nutzers, führt die mangelnde Passung zu bestehenden Kommunikationsstrategien des Least Collaborative Effort-Modells auch zu spezifischen **Bedienfehlern**. So animiert insbesondere eine überflüssig erscheinende Nachfrage zu einer zu frühen Spracheingabe. Durch überlappende Äußerungen versuchen Nutzer im Rahmen ihrer Möglichkeiten die Interaktion zu beschleunigen, stoßen damit allerdings an die Effektivitätsgrenzen des Systems. Damit entstehen wiederrum Fehler- und Reparaturkosten, die durch eine effizientere Gestaltung des SDS behoben werden könnten.

In einer WoZ-Studie mit einem sprachgesteuerten Telefonagenten setzten Brennan und Hulteen (1995) ein konservatives Grounding Criterion mit statischem Dialogverhalten um. Ähnlich wie aktuelle SDS, meldete dieses konstant den Zustand 3 (z. B. *Ich habe verstanden, dass Sie Susan anrufen wollen, ist das korrekt?*) und den Zustand 6 (z. B. Wähltöne) zurück. Sie stellten dabei fest, dass viele Nutzer von der expliziten Bestätigungsanfrage genervt waren. Jene Dialogschritte wurden nur dann akzeptiert, wenn sie durch Fehlerkennungen im Vorfeld legitimiert werden konnten.

Maciej und Vollrath (2009) konnten darüber hinaus belegen, dass der Dialogfluss Einfluss auf die Fahrleistung haben kann. So führte eine einzelne Bestätigungsaufforderung am Dialogende zu 15% weniger Spurabweichung als das multiple, schrittweise Nachfragen. Damit konnte ein starker Ablenkungseffekt durch die manuelle Bestätigung des Inputs nachgewiesen werden.

Es kann geschlossen werden, dass SDS die Reduktion des Gesamtaufwands der Kommunikation mehrheitlich dem Nutzer überlassen. Die geringe Dialogeffizienz, die in einem inflexiblen statischen GC begründet liegt, widerspricht dem Ansatz des LCE und führt zu einer geringen nutzerseitigen Effizienzbeurteilung. Neben der Unzufriedenheit über die redundanten Dialogschritte, wirken sich die expliziten Nachfragen auch negativ auf die Systemkompetenzbeurteilung der Nutzer aus. Die expliziten Bestätigungsanfragen begünstigen überlappende Spracheingaben, die die Effektivität des Systems und die Zufriedenheitsbeurteilungen der Nutzer gefährden. Somit werden Sprachdialogsysteme auch in jenem Aspekt ihrer Verantwortung als kollaborativer Empfänger nicht gerecht. Im Folgenden sei das letzte adressierte Element des Groundings, das Alignment, im Hinblick auf die Umsetzung in aktuellen Systemen beschrieben.

4.1.3 Alignment automotiver SDS

Obwohl Menschen ein hohes Ausmaß an Alignment in zwischenmenschlichen Dialogen zeigen (Kap. 3.2.3), verwenden alle Sprachdialogsysteme im Fahrzeugkontext **konstante Systemreaktionen**. Das folgende Dialogbeispiel verdeutlicht, das keine linguistische Angleichung an die Wortwahl des Nutzers stattfindet.

Nutzer	Verbindungsaufbau zu Helena Otto privat.
System	Möchten Sie Helena Otto zu Hause anrufen?

Neben der Funktion als Verständnisevidenz, können Vorteile der systemseitigen Anpassung auch aus den Ursachen des zwischenmenschlichen Alignments abgeleitet werden. Im Folgenden werden aus diesen Vorteilen die möglichen Folgen fehlenden Alignments antizipiert.

Folgen mangelnden Alignments

Im Rahmen der **vermittelten Prozesse** des Alignments, kann die menschliche Fähigkeit, das kommunikative Verhalten an den Partner anzugleichen, als Schlüssel zur erfolgreichen Kommunikation bezeichnet werden (Goleman, 2006). Verwendet das System nicht den vom Nutzer eingeführten Begriff, so kann dies als ein Hinweis auf eine mangelnde Passung zwischen Intention und Verständnis nutzerseitig interpretiert werden. Dies erschwert den Aufbau eines Common Grounds und führt im schlimmsten Fall zu **Missverständnissen**, die nicht aufgelöst werden können. So berichtete Hassel (2006) von endlosen Korrekturschleifen in Dialogen, da das Sprachdialogsystem auf die Nutzereingabe *Verkehrsinfo ein* immer mit *Verkehrsfunk ein.* reagierte. Da einzelne Nutzer dies nicht als die gleiche semantische Kategorie interpretierten, versuchten sie das System zu korrigieren und endeten in dialogischen Sackgassen.

In Kapitel 3.2.3 wurde argumentiert, dass durch die **Prozesse des Primings** eine Kopplung der Sprachproduktion und des -verstehens erreicht wird, die die Dialogführung auf mentaler Ebene stark vereinfacht (Pickering & Garrod, 2004). Zeigt ein SDS nun keine Anpassung an die Eingabe des Nutzers, so kann davon ausgegangen werden, dass der kognitive Verarbeitungsaufwand auf Seiten des Nutzers steigt. Im Bereich des Groundings können also höhere Kosten hinsichtlich der Rezeption und des Verständnisses auftreten, als man dies aus gewöhnlichen zwischenmenschlichen Dialogen kennt.

Bezogen auf die **strategischen Aspekte** des Alignments konnte nachgewiesen werden, dass im zwischenmenschlichen Dialog die Sprecher positiver beurteil wurden, die eine Angleichung des Wortschatzes zeigten (Bradac et al., 1988; Branigan et al., 2010). Sprachdialogsysteme, die bisher konstant an ihrer Wortwahl festhalten, könnten demnach als unhöflich oder korrigierend empfunden werden und bei den Nutzern negative Affekte auslösen. Aus zwischenmenschlichen Dialogen bringen Nutzer eine starke Erwartungshaltung mit, dass akzeptierte Begriffe auch beim Gesprächspartner Verwendung finden (Metzing & Brennan, 2003). Sprachdialogsysteme, die keine Angleichung an den Nutzer bieten, laufen demnach Gefahr, als weniger transparent und sympathisch wahrgenommen zu werden.

Es kann geschlussfolgert werden, dass SDS bisher eine unzureichende Anpassung an die Nutzer oder die Situation zeigen und ihrer Verantwortung als Empfänger nicht vollständig erfüllen. Um die Aufwandskosten des Fahrers bei der Sprachbedienung zu reduzieren und Rahmenbedingungen für einen nutzerfreundlichen Dialog zu schaffen, soll bei der Optimierung der SDS die Erleichterung der Grounding-Prozesse im Fokus stehen. Das folgende Kapitel soll dazu genutzt werden, um zu belegen, dass es erfolgsversprechend erscheint, die drei Elemente des Groundings in Angleichung an den zwischenmenschlichen Dialog auch bei dem Design von SDS zu berücksichtigen.

4.2 Übertragung der Konversationsmaximen

Anhand von Wizard-of-Oz–Experimenten konnte belegt werden, dass Nutzer nahezu deckungsgleiche prozeduralen Ziele und Erwartungshaltungen im Dialog mit einer Maschine zeigen. Bernsen und Dybkjaer (2004) erarbeiteten Gestaltungsrichtlinien für nutzerfreundliche SDS, die große Schnittmengen mit den Maximen zwischenmenschlicher Kommunikation von Grice (1975) aufweisen (Kapitel 3.1.2.1).

Aktuelle Studien mit dem Dialogsystem von Bernsen und Dybkjaer (2004), welches nach diesen Gestaltungsrichtlinien erstellt wurde, konnten zeigen, dass durch die Beachtung dieser umfassenden kooperativen Kommunikationsmaximen die Nutzerfreundlichkeit der SDS erhöht werden

konnte. Gleichzeitig stellt eine starke Verletzung der Erwartungshaltungen der Nutzer potenzielle Fehlerquellen dar.

Es kann demnach geschlossen werden, dass die prozeduralen Ziele der Collaborative Theory auch bei der Interaktion mit SDS Gültigkeit besitzen. Im Rahmen der zielgerichteten Tätigkeit des Informationsaustauschs müssen Mensch und Maschine kooperieren, damit ihre Kommunikation erfolgreich verläuft. Beide müssen gewisse Konversationsmaximen einhalten, um nicht gegen das Kooperationsprinzip zu verstoßen. In den folgenden Kapiteln soll die konkrete Umsetzung argumentiert und Untersuchungen zitiert werden, deren Befunde die Idee der Übertragung der zwischenmenschlichen Grounding-Elemente stützen.

4.2.1 Übertragung Feedback

Wie Kapitel 4.1.1 belegen konnte, sind dem Nutzer bisher weder der Systemstatus, noch die geteilte Wissensbasis mit dem System transparent. Um eine kooperative Gesprächsführung mit einem SDS zu erreichen, sollte der bisher leere Raum zwischen der nutzerseitigen Eingabe und der Aktionsankündigung (Zustand 5 des adaptierten Contribution Models) durch sofortige Empfangsbestätigungen und dialogstrukturierende Signale gefüllt werden. Somit sollte kontextsensitives Feedback jederzeit kombinierte Rückmeldungen darüber geben, in welchem Dialogstatus sich das System befindet und welche Dialoginhalte bereits erkannt wurden. Damit würden nicht nur die Konversationsmaximen eingehalten, sondern auch der Anspruch des LCE erfüllt und das Grounding mit einem SDS erleichtert.

Bezugnehmend auf die zwei **Dialogspuren** nach Clark (1996) ist Sprache für diesen Zweck als alleinige Interaktionsmodalität nur bedingt hinreichend. Neben der Inhaltsebene, füllen vor allem paralinguistische Signale in der zwischenmenschlichen Kommunikation die zweite Dialogspur. Acknowledgements (vgl. Kap. 3.2.1), die dem Sprecher Verständnisevidenzen liefern oder der Disambiguierung dienen, nutzen verschiedene sensorische Kanäle. Sie können beispielsweise durch Gestik, Mimik oder paraverbale Geräusche (z. B. Backchannels, vgl. Kap. 3.2.1) ausgedrückt werden. Für die Übertragung zwischenmenschlicher Feedbackstrategien auf den Mensch-Maschine-Dialog soll der visuelle Kanal genutzt werden.

Dafür spricht, dass **visuelles Feedback** auch in der zwischenmenschlichen Kommunikation eine große Rolle spielt, um die zweite Dialogspur, die primär die Grounding- und Empfangshandlungen beinhaltet, zurückzumelden. Nahezu jede sprachliche Interaktion mit gegenseitiger Sichtbarkeit wird von visuellen Stimuli, wie Gesten und Mimik begleitet (Gibbon, 2000). Dabei kann die zusätzliche Modalität die Kommunikation nicht nur effektiver gestalten. Im Rahmen von Analysen zwischenmenschlicher Dialoge konnte belegt werden, dass die Hinzunahme des visuellen Kanals aufgabenorientierte (zwischenmenschliche) Dialoge beschleunigen kann (Gergle et al. ,2004;

Clark & Krych, 2004). Die zusätzlichen Informationen verringern den geteilten Aufwand und beeinflussen das Grounding positiv. So sind die visuellen Evidenzen (z. B. Kopfnicken) häufig ausreichend um Akzeptanz auszudrücken und auf eine ausführliche (verbale) Bestätigung kann somit verzichtet werden. Clark und Krych (2004) schließen, dass einige Gesten effizientere Grounding-Strategien bieten, als rein sprachliche Äußerungen.

"At least some gestural acts, we assume, are more efficient for grounding than the vocal acts that would be needed without them." (ebd., S. 67)

Da der Empfänger in der zwischenmenschlichen Kommunikation sich häufig visueller Reize (Gestik, Mimik) bedient, um dem Sprecher seinen Zustand zu vermitteln, sollten diese auch in Interaktion mit einem SDS vermittelt werden. Sowohl die Ergebnisse der Studie zur Häufigkeit des TTE-Fehlers (Kapitel 2.3), als auch die Ergebnisse von Kainer (2007) zeigen, dass Versuchspersonen eine visuelle Reaktion des Systems erwarten und aktiv nach ihr suchen.

Das Vorgehen, die Systemzustände und Dialoginhalte grafisch zu vermitteln, bestätigen auch Bernsen und Dybkjaer (2001), die das multimodale Feedback als eines von 14 spezifischen Einflusskriterien auf die Gebrauchstauglichkeit von Sprachdialogsystemen identifizieren konnten. Sie kommen zu dem Schluss, dass SDS davon profitieren würden, wenn man abseits der sprachlichen Schnittstelle auch andere Feedbackkanäle motiviert.

„*Telling the user is not always good enough […] It may therefore be a good thing for SLDSs to provide several different kinds of feedback to their users. "* (ebd., S. 14)

Die ergänzenden Visualisierungen der Systemzustände und Dialoginhalte würden die Möglichkeit bieten, verschiedene **Einschränkungen** der Modalität Sprache zu kompensieren (vgl. Kapitel 2.3).

- **Persistenz-Problem:** Im Gegensatz zur Sprachausgabe, die nur temporär wahrnehmbar ist (vgl., Morgan et al., 2001, S. 11), sind visuelle Rückmeldungen auf dem Display zeitpersistent. Den Nutzern bietet dies höhere Freiheitsgrade in der Entscheidung, wann sie Inhalten der Zweitaufgabe Aufmerksamkeit zuwenden. Auch wenn Nutzer einer akustischen Nachfrage des Systems aufgrund eines komplexen Verkehrsmanövers keine Beachtung schenken können, so können sie beispielsweise durch die Anzeige der letzten Systemausgabe zu einem weniger beanspruchenden Moment wieder mit dem Dialog fortfahren. Somit kann die Dauerhaftigkeit der visuellen Repräsentation subjektiv als Entlastung empfunden werden.
- **Conversational Model:** Das visuelle Feedback bietet weiterhin die Möglichkeit, Sprachdialogsysteme zum Pol des "**keypad models**" (Kap. 2.3.2) zu verschieben und damit ihrer Intuitivität zu erhöhen. Durch eine transparente Darstellung der Inhalte und Prozesse können

dem Nutzer alle Handlungsoptionen klar aufzeigt werden. In Bezug auf Clarks Collaborative Theory (1996) bietet gerade die Möglichkeit, die Dialoginhalte visuell darzustellen, die Gelegenheit den Common Ground als externe Repräsentation bereitzustellen. Somit können nicht nur Missverständnisse vermieden, sondern der geteilte Kontext auch im positiven Sinne sichtbar gemacht werden. Die Anzeige der Dialoginhalte bietet neben einer frühzeitigen Möglichkeit zur Korrektur auch eine Memorierungshilfe und kann dem Nutzer das effektive Bedienen des SDS beibringen.

- **Informationsbandbreite:** Sprachausgaben können in der Regel nur **sequentiell** präsentiert werden (Schmandt, 1994, S. 102; Morgan et al., 2001, S. 11) und sind damit kognitiv anstrengend und zeitaufwendig. Während über den sprachlichen Kanal so nur eine Wort-für-Wort-Darbietung realisiert werden kann, ist durch eine visuelle Darbietung das Erfassen mehrere Stimuli oder einer kompletten Sinneinheit mit einem Blick möglich. Gerade bei der Darbietung umfangreicher akustischer Informationen kann die zeitliche Dauer der mentalen Ablenkung vom Fahrgeschehen durch eine visuelle Darbietung sogar reduziert werden, da Menschen sehr effizient visuelle Informationen verarbeiten können[15]. Es kann davon ausgegangen werden, dass bei einer Systemantwort in Form einer alleinigen Sprachausgabe ohne Kontextinformationen und mit Fahrgeräuschen eine erhöhte mentale Beanspruchung bei der Dekodierung vorliegt. Die Ergänzung der Sprachausgabe um textuelle oder zeichenartige Bestandteile, die auf eindeutigen und schnell erfassbaren Schriftzeichen beruhen, bietet somit Vorteile bei der Dekodierung.
- **Simultanität:** Insbesondere Meta-Informationen, wie die zu erwartende Verarbeitungsdauer des Systems, lassen sich visuell sehr viel verständlicher darbieten. Ebenso können Meta-Informationen, wie der Empfangsstatus (Zustand 2), über den visuellen Kanal parallel zur Sprachproduktion des Nutzers dargeboten werden, ohne diesen zu unterbrechen. Die empfängerseitigen Rückmeldungen lassen sich somit durch Visualisierungen sehr viel leichter umsetzen.

Für die Übertragung der Feedbackstrategien mit Hilfe des visuellen Sinneskanals sprach neben diesen Vorteilen auch, dass sich alternative Gestaltungsansätze noch auf keinem zufriedenstellenden technischen Niveau befinden. So wurde sich beispielsweise gegen die akustische Imitation von **Backchannels** oder dialogstrukturierenden Signalen durch das System entschieden, da eine technische Prädiktion, wann eine solche Äußerung erfolgen sollte, noch auf keinem zuver-

[15] Beim Sprechen werden beispielsweise durchschnittlich 110 bis 120 Worte pro Minute produziert, während beim Lesen ca. 300 Worte pro Minute verarbeitet werden können (Kozma, 1991).

lässigen Level möglich ist. Die bisher maximal erzielte Treffsicherheit von Cathcart et al. (2003) von 35 % für die englische Sprache, kann nicht als zufriedenstellend bewertet werden.

Aufgrund ihrer schlechten Differenzierbarkeit und antizipierten Interferenzeffekten wurde sich gegen die Alternative, **akustische (nonverbale) Signale** wie Spearcons[16], Earcons[17] oder Auditory Icons[18] (Brewster, 2002) zur Vermittlung des Systemstatus zu verwenden, entschieden. So ist der akustische Kanal bereits durch die Aufnahme von Inhaltsinformationen des Dialogs belegt. Laut des Multiplen Ressourcen Modells (Kapitel 2.3.1) könnte eine zusätzliche akustische Zustandsdarbietung nicht nur irritierend, sondern auch ablenkend sein. Darüber hinaus ist ein akustisches Signal ähnlich wie Sprache vergänglich und kann somit nur kurz dargeboten werden. Prozessbegleitende Signale, wie die des Empfangens oder Verarbeitens, ließen sich damit nur schwer abbilden. Earcons, Spearcons oder Auditory Icons während der Spracheingabe des Nutzers (zur Signalisierung des Zuhörens) oder während der Verarbeitungsphase wären demnach nur bedingt geeignet, um Dialoge ressourcenschonender zu gestalten. Auch verbale Rückmeldung der Zustände wie beispielsweise die Ankündigungen *Ich höre, Aha, Ich verarbeite* wären an dieser Stelle mit denselben Herausforderungen wie die nonverbalen Signale konfrontiert.

Bei der Übertragung der Feedbackprozesse mittels visueller Rückmeldungen lassen sich allerdings auch **negative Effekte** antizipieren. Laut des Multiplen Ressourcen Modells (Kapitel 2.3.1) ließen sich Interferenzen des visuellen Feedbacks der Tertiäraufgabe mit der Primärraufgabe (Fahren) vermuten. Um die Vorteile der Sprachbedienung im automotiven Kontext beizubehalten, sollen umfassende Strategien zur Vermeidung von fahrerbezogener Ablenkung Umsetzung finden. Dabei soll das visuelle Feedback sehr abstrakt und einfach gestaltet werden. Die Umset-

---

[16] **Spearcons:** extrem schnell abgespielte Sprachausgaben, die ohne zusätzliche Kontextinformationen oder Vorwissen nicht verstanden werden können

[17] **Earcons**: kurze spezifische Tonfolgen oder musikalische Konzepte (Blattner et al.,1989), die sich durch Frequenz, Klangfarbe, Dauer und Tempo unterscheiden und deren zugrundeliegende Semantik durch den Nutzer erst erlernt werden muss

[18] **Auditory Icons**: direkte oder metaphorische Repräsentation eines semantischen Konzepts (Gaver, 1986; z.B. Geräusch von zerknülltem Papier wenn man eine Datei in den Papierkorb eines Macs verschiebt)

zung, die in Kapitel 6.1.2.2 detaillierter dargestellt wird, soll auf eine schnelle Wahrnehmbarkeit der Zustände und Inhalte ausgelegt werden und so nur geringe Blickabwendungen hervorrufen.

Einerseits soll die akustische Ausgabe aufgrund der Anforderungen der Verkehrssituation stets die **Hauptmodalität** der Interaktion bleiben. Die sprachlichen Ausgaben des Systems sollen zu jeder Zeit eine hinreichende Informationsquelle darstellen, um das Interaktionsziel zu erlangen. Als Gestaltungsmaxime der Visualisierungen gilt somit, dass ein Dialog auch ohne Blick auf die grafische Anzeige erfolgreich beendet werden kann. Da die visuellen Rückmeldungen somit nur eine unterstützende Funktion erfüllen, bedeutet dies übersetzt in das in Kap. 2.1.1 eingeführte vereinfachte Klassifikationsschema multimodaler Systeme nach Nigay und Coutaz (1993), dass eine parallel-komplementäre Interaktion für das Zustandsfeedback und eine parallel-redundante Interaktion für das Inhaltsfeedback angestrebt wurde. Während die Rückmeldung der Systemzustände lediglich über die grafische Modalität erfolgt und damit verschiedene Informationen über unterschiedliche Modalitäten parallel verarbeitet werden, erfolgt die Rückmeldung der Dialoginhalte parallel über verschiedene Modalitäten. Angrenzende Studien zu multimodalen Bedienkonzepten im Fahrzeug konnten diesbezüglich belegen, dass kombinierte Stimuli mit der gleicher Semantik zu verstärkter Aufmerksamkeit und kürzeren Reaktionszeiten führen (Selcon & Taylor, 1995). Auch Bengler (1995) konnte beweisen, dass Navigationssysteme mit kombinierten visuell-akustischem Feedback eine geringere mentale Beanspruchung hervorrufen. In beiden Untersuchungen stellte die akustische Ausgabe jeweils die primäre Informationsquelle dar, die durch visuelle Informationen ergänzt wurde.

Da die visuellen Anzeigen nicht zwingend benötigt werden, sondern die Dialogführung lediglich unterstützen, kann davon ausgegangen werden, dass Nutzer ihr Interaktionsverhalten an die entsprechende Fahrtsituation anpassen und **kompensatorische Maßnahmen** vornehmen. Grundsätzlich sei dabei auf die Fähigkeit der Fahrer verwiesen, ihre Aufmerksamkeitsverteilung an die Gegebenheiten der primären Fahraufgabe und die Eigenschaften der Zweitaufgabe anzupassen. Studienergebnisse zeigen, dass Nutzer Blickabwendungen im Fahrtkontext in der Regel stark kontrolliert vornehmen. Es konnten Blickspannen zwischen 0.6 und 1.7 Sekunden festgestellt werden, wobei im Regelfall die Dauer von einer Sekunde nicht überschritten wird (Wierwille et al., 1991; Wikman et al., 1998). Reicht diese Blickdauer nicht aus, so kompensieren Nutzer dies mit erhöhter Blickfrequenz.

Dass die Anzeige von zusätzlichen Informationen sich nicht zwingend negativ auf die Fahraufgabe auswirken muss und die subjektive Bewertung dadurch sogar noch gesteigert werden kann, belegen Curin et al. (2011) mit einer automotiven Sprachapplikation, die das Diktieren und Korrigieren von Textnachrichten während der Fahrt ermöglicht. Dabei wurden drei Umsetzungen diskutiert: Die Erste ermöglichte eine Interaktion ohne Blickabwendung, da auf dem Display nichts angezeigt wurde (*eyes-free*). Die zweite Variante zeigte den diktierten Text vollständig an (*full*

view), während die dritte Variante nur den aktiven Text anzeigte, der gerade von der TTS vorgelesen wurde (*strip view*). Auch wenn die *eyes-free*-Variante die geringste Spurablage, verglichen zu den anderen beiden Varianten, zeigte[19], so wurde sie von den Versuchspersonen am höchsten bei dem Kriterium der mentalen Belastung eingeschätzt. So wurde es als stark beanspruchend erlebt, nicht zusätzlich zur Sprachausgabe zu sehen, was von dem System erkannt wurde. Obwohl eine größere Blickabwendung beim *strip view* beobachtet werden konnte, erlangte dieser eine bessere Gebrauchstauglichkeitseinschätzung als die *eyes-free*-Variante. Eine ergänzende visuelle Anzeige scheint also in diesem Fall die mentale Beanspruchung zu reduzieren, wenngleich die Augen nicht permanent auf der Straße sind.

Nach der Umsetzungsbeschreibung der Übertragung der Feedbackstrukturen zwischenmenschlicher Dialoge, soll im Folgenden auf die Übertragung von dynamischem Rückfrageverhalten eingegangen werden.

4.2.2 Übertragung Grounding Criterion

Wie bereits im Kapitel 4.1.2 beschrieben, liefern bisherige Sprachdialogsysteme Verständnisevidenzen immer explizit in Form von Wiederholungen oder Bestätigungsanfragen des Nutzerinputs nach jeder Eingabe (Brennan & Hulteen, 1995). Eine derartige „*forcing function*“ (Norman, zitiert nach Brennan, 1998, S. 210) hilft zwar Erkennungsfehler zu verhindern und dient der Stabilisierung des Dialogs, kann aber auch unnötig und ermüdend sein. Denn einerseits werden so die systemseitigen Interpretationen der Nutzereingabe informativ zurückgemeldet und Reparaturen ermöglicht, aber andererseits erhöht dieses statische Dialogverhalten den nutzerseitigen Gesprächsaufwand.

Brennan und Hulteen (1995) führen daher das „**System Grounding Criterion**“ (SGC) ein, um eine intelligente Anpassung im Dialogverhalten von Systemen, ähnlich der zwischenmenschlichen Kommunikation, zu ermöglichen. Im Rahmen des adaptierten Contribution Modells (vgl. Kap. 4.1.1) kann das System für jeden Zustand, in dem es sich befindet, explizite oder implizite Evidenz vom Nutzer einfordern. Das SGC bestimmt dabei, äquivalent zum Grounding Criterion der zwischenmenschlichen Kommunikation, wie umfangreich die vom System geforderte Evidenz ausfällt. Da jeder redundante Dialogschritt den kollaborativen Aufwand erhöht (Clark & Brennan, 1991), entscheidet das System dynamisch, ob eine explizite Evidenz, z. B. in Form einer Bestäti-

[19] Betrachtet man die Spurabweichung ohne die durch den LCT evozierten Spurwechselmanöver, so kann zwischen den drei Varianten keine Differenz mehr festgestellt werden.

gungsaufforderung, oder eine implizite Evidenz durch die Übernahme des nächsten relevanten Dialogschritts gerechtfertigt ist.

Im Rahmen des Ansatzes von Brennan und Hulteen (1995) determinieren drei **Parameter** die Höhe des SGC und damit welche Menge an Feedback, innerhalb welcher Dialogzustände geäußert wird: die Dialoghistorie, die Bedienumgebung und das Aufgabenmodell. Mussten im Laufe des Dialogs bereits Korrekturen vorgenommen werden, sind Störgeräusche präsent oder handelt es sich um eine kritische Aufgabe, so wird das SGC konservativ (hoch) gewählt. Sind diese Risikofaktoren allerdings nicht gegeben, so kann auf die Bestätigungsanfragen verzichtet werden und das SGC liberal (niedrig) gewählt werden. Daraus folgt beispielsweise, dass in einer Umgebung mit vielen Störgeräuschen eher explizite Evidenzen für niedrige Empfängerzustände präsentiert werden. Wohingegen bei unkritischen Aufgaben, deren Missverstehen keine schweren Konsequenzen hätte, ein liberales SGC ausreicht und Evidenzen nur implizit erbracht werden.

Damit entsteht eine hohe Flexibilität des Interaktionsablaufs, die zu einer gewissen Systeminkonsistenz führen kann. Je nach Dialoghistorie oder Bedienumgebung kann ein Dialogziel eine unterschiedliche Anzahl an Interaktionsschritten bedingen. Die Dynamik des Dialogs kann somit zu unterschiedlichen Dialogabläufen führen, die dem Nutzer nicht zwangsläufig eingängig sind und damit die Erwartungen an die Konsistenz des Systems verletzen. Im Folgenden wird dies als **Konsistenzdilemma** bezeichnet, da hier die Regeln der Schnittstellengestaltung den zwischenmenschlichen Verhaltensweisen entgegenstehen. Während von Maschinen immer erwartet wird konsistent zu reagieren, zeigen Menschen selbst in zielorientierten Dialogen eine hohe Variabilität.

Bezüglich der nutzerseitigen Akzeptanz lassen sich in der Literatur jedoch keine negativen Effekte der dynamischen Reaktionen eines SDS feststellen. So konnten Brennan und Hulteen (1995) in einer WoZ-Studie sogar belegen, dass viele Nutzer die expliziten Bestätigungsanfragen nur dann akzeptierten, wenn es im Vorfeld zu Fehlerkennungen (Substitutionsfehlern) gekommen war. Ferner evaluierten Chu-Carroll und Nickerson (2000) ein System, das die Initiative der Dialogführung automatisch anpassen konnte und belegten, dass das adaptive System im Hinblick auf die Effizienz und Usability besser bewertet wurde als das konstant-reagierende System.

Auch Litman und Pan (2002) implementierten ein System, das die Dialogstrategie je nach Gesprächsverlauf anpasst. Zunächst zeigt Ihr **TOOT-System** eine aggressive Interpretation der Nutzereingabe. Erst wenn eine hohe Anzahl an Fehlerkennungen detektiert wurde, wird die Rückmeldestrategie des Systems angepasst und die Dialoge ausführlicher gestaltet. Litman und Pan (2002) konnten durch diese Abwärtskorrektur die Dialoge effektiver und nutzerfreundlicher gestalten.

Die Befundlage verdeutlicht, dass es vielversprechend ist, die Interpretationsstrategie des SDS flexibel zu gestalten und das SGC, welches das Ausmaß der Evidenzen definiert, an die Dialogsituation zu knüpfen. Allerdings beziehen sich die vorgestellten Arbeiten primär auf Systeme mit Kenntnis des aktuellen Nutzers oder auf Systeme, denen mehrere Dialogschritte zur Anpassung zur Verfügung stehen (ebd.). In Kapitel 6.2.2.1.1 wird eine Umsetzung für automotive SDS vorgeschlagen, die dynamisches Rückfrageverhalten auch mit adäquaten technischem Aufwand ermöglicht. Dabei wird die **Erkennerkonfidenz** der letzten Nutzereingabe als aktueller Schätzer der Systemsicherheit herangezogen.

Nach der Diskussion einer Flexibilisierung des systemseitigen Rückfrageverhaltens, soll im Folgenden auf die Angleichungsprozesse im Mensch-Maschine-Dialog eingegangen werden.

4.2.3 Übertragung Alignment

Vergleichbar zur Bidirektionalität der Feedbackprozesse, können auch zwei Richtungen des Alignments identifiziert werden. Anfänglich wird ein empirischer Überblick darüber gegeben, wie und in welchem Umfang Menschen ihre Äußerungen an Systeme anpassen. Abschließend werden Effekte einer systemseitigen Angleichung diskutiert.

In einer Vielzahl an Untersuchungen konnte belegt werden, dass Menschen sich auf verschiedenen Ebenen der linguistischen Repräsentation an Maschinen anpassen. So konnte Bell et al. (2003) nachweisen, dass Sprecher ihre **Sprechrate** an Computer angleichen. Dabei waren sie sensitiv gegenüber Feedback und passten ihr Ausmaß an Alignment an, wenn der Computer mit zu schnellen Eingaben Probleme zeigte. In einer Reihe von Untersuchungen mit Kindern konnten Oviatt et al. (2004) nachweisen, dass sie sich ihrem animierten Gesprächspartner konsistent auf mehreren **akustisch-prosodischen Ebenen** (Amplitude, Pausenstruktur, etc.) anglichen. Aktuelle Studien weisen diesen Effekt auch bei erwachsenen Nutzern nach (Suzuki & Katagiri, 2007). Brennan (1996) untersuchte mit Hilfe eines WoZ-Szenarios **lexikalisches Alignment** in Mensch-Maschine-Dialogen. Sie fand heraus, dass Sprecher sich an die Wortwahl des Computers anglichen, nachdem der Computer ihre Wortwahl hinterfragt hatte. Auch Richards und Underwood (1984) konnten den Einfluss der Äußerungen des Systems auf die der Probanden nachweisen. Je ausführlicher und höflicher die Eingabeaufforderung ausfiel, umso wortreicher waren die Nutzeräußerungen. Branigan et al. (2010, 2003) konnten in diversen Studien belegen, dass sich Nutzer stark an die **Satzstruktur**, die vom Computer verwendet wird, anpassen.

Vergleichende Untersuchungen weisen darauf hin, dass das Ausmaß der Angleichung im Mensch-Maschine-Dialog größer ist, als im Mensch-Mensch-Dialog. So traten beispielsweise syntaktische Angleichungsprozesse in Branigan et al. (2010, 2003) Studien stärker in Mensch-Maschine-Dialogen auf als in zwischenmenschlichen Dialogen. Diese Befundlage wurde durch

ein weiteres Experiment bezüglich des **lexikalischen Alignments** bestätigt (Branigan et al., 2004 zitiert nach Branigan et al., 2010, S. 2365). Dabei wurden in einer Vorstudie für verschiedene Objekte präferierte und weniger präferierte Bezeichnungen identifiziert. Konfrontiert mit einem menschlichen Gesprächspartner passten sich die Versuchspersonen nur in 15% der Fälle an eine Bezeichnung an, die sie nicht präferierten. Glaubten sie sich allerdings in einer Interaktion mit einem Computer, glichen sie sich zu 67% an diese Bezeichnung an. Bemerkenswert ist der Fakt, dass Menschen das Vokabular des Computers auch dann nutzen, wenn sie selbst ein anderes Wort bevorzugten, um den Dialog zu erleichtern. Auch Pearson et al. (2006) konnten eine Anpassung an das **Empfängermodell** eindrucksvoll nachweisen, indem den Versuchspersonen zwei unterschiedliche Startscreens nach dem Hochfahren des Dialogsystems präsentiert wurden. Ein Screen zeigt die Bezeichnung *Basic Version*, der andere *Advanced Version*. Obwohl es sich funktional um das gleiche System handelte, schätzten Versuchspersonen, dass *Advanced* System kompetenter ein als das *Basic* System. Hinsichtlich beider Varianten zeigten sie eine lexikalische Anpassung. Das Ausmaß des Alignments war allerdings gegenüber dem Basissystem sehr viel größer. Die erhöhte Anpassung an einen maschinellen Gesprächspartner deutet somit auf eine große Bedeutung der **vermittelten Prozesse** in der asymmetrischen Beziehung zwischen Mensch und Maschine hin. Der höhere Ressourcenaufwand im Vergleich zu zwischenmenschlichen Dialogen wird dabei vom Nutzer in Kauf genommen, um die Erfolgswahrscheinlichkeit der Kommunikation zu erhöhen.

Obwohl Menschen sich in einem hohen Ausmaß an die Systeme anpassen, sind bisher nur wenige Systeme umgesetzt worden, die sich der Nutzereingabe angleichen. Dementsprechend haben sich nur wenige Studien mit der Frage beschäftig, wie sich Alignment seitens des Systems auf den Nutzer auswirkt. Es gibt vereinzelt Nachweise, dass eine Angleichung durch den Computer **positive Affekte** beim Nutzer auslöst. Hinsichtlich non-linguistischer Anpassung konnten Bailenson und Yee (2005) berichten, dass Nutzer einen animierten Computeragenten positiver bewerteten, wenn dieser ihre Kopfbewegungen imitierte. Hat sich ein sprachbedienbares Telefonsystem an die Sprechrate der Nutzer angepasst, so bewerteten sie auch dieses besser als ein System, dass diese Angleichung nicht vornahm (Ward & Nakagawa, 2002). An dieser Stelle sei auch auf Weizenbaums Eliza (1966) verwiesen, die durch ein Spiegeln der Nutzereingaben ganze Gespräche aufrecht erhalten und positive Emotionen bei ihren Nutzern auslösen konnte. Ein Dialog, in dem Eliza das Gespräch durch Angleichungsprozesse aufrechterhält, kann wie folgt aussehen:

Nutzer	Mein Freund hat mich hergeschickt.
Eliza	Ach so, Ihr Freund hat Sie hergeschickt.
Nutzer	Er sagt, ich sei oft depressiv.
Eliza	Tut mir leid zu hören, dass Sie oft depressiv sind.

Es konnte ebenso gezeigt werden, dass auch die Formulierungen des Systems einen starken Einfluss auf die Systembewertung hatten (vgl. Bell et al., 2003; Nenkova et al., 2008; Porzel, 2006). So konnten Porzel et al. (2006) nachweisen, dass die Dialogeffizienz und Nutzerzufriedenheit durch eine Angleichung des Dialogsystems an die Bezeichnungen des Nutzers erhöht werden konnte. Insbesondere Novizen profitierten in seinem Problemlöseszenario davon, dass sich das erklärende System an ihre Formulierungen anpasste.

Die detaillierte Darstellung der Gestaltungsprobleme aktueller SDS zeigt auf, dass sie dem kollaborativen Charakter der Kommunikation noch nicht gerecht werden. Durch die Übertragung der drei Grounding-Elemente können, bezugnehmend auf die dargestellte Befundlage, allerdings positive Effekte auf die Gebrauchstauglichkeit erwartet werden.

4.3 Fazit

Anhand der Collaborative Theory wurde in Kapitel 4 eine Statusanalyse aktueller Sprachdialogsysteme durchgeführt. Dabei wurde festgestellt, dass die Aufwandsminimierung und die Verantwortung für einen erfolgreichen Abschluss der Kommunikation hauptsächlich beim Nutzer liegen (Brennan, 1998). Es lassen sich zusammenfassend dargestellt drei Problembereiche identifizieren, in denen aktuelle Sprachdialogsysteme von den zwischenmenschlichen Kommunikationsstrategien abweichen und damit Konversationsmaximen und Erwartungshaltungen der Nutzer nicht erfüllen:

(1) Es mangelt Sprachdialogsystemen an **Feedback**, welches dem Nutzer transparent den Status des Systems vermittelt und damit seiner Verantwortung als Empfänger nachkommt. Auch dialogstrukturierende Signale sind nicht ausreichend vorhanden, was zu einer großen nutzerseitigen Unsicherheit führt, die nicht selten in Bedienfehlern mündet.
(2) Das Bestreben den **Least Collaborative Effort** zu erreichen wird gestört. Durch unflexible Interaktionsstrukturen und viele Bestätigungsaufforderungen ist der nutzerseitige Aufwand für den gegebenen Zweck der Kommunikation zu groß. Dieser Mangel an Effizienz führt nicht nur zur nutzerseitigen Frustration, sondern provoziert Bedienfehler und ein geringes Zuverlässigkeitsempfinden.
(3) Systemausgaben werden derzeit nicht an die Nutzereingaben angeglichen. Damit werden die Erwartungshaltungen der Nutzer bzgl. des **Alignments** nicht erfüllt und die Sprachbedienung gestaltet sich anstrengender als das Vorbild des nahezu ressourcenfrei ablaufenden Dialogs zwischen Menschen.

Die Übertragung von zwischenmenschlichen Kommunikationsstrategien auf Sprachdialogsysteme wurde als Lösungen dieses Gestaltungsproblems im zweiten Abschnitt des Kapitels diskutiert.

Verschiedene Forschungsbefunde wurden herangezogen, um den Erfolg der Übertragung kollaborativer Kommunikationsprinzipien auf den Mensch-Maschine-Dialog zu belegen.

Im Konkreten wurde sich bei dem Transfer zwischenmenschlicher Rückmeldeprozesse für die Darbietung der Systemzustände und Dialoginhalte entschieden. Nach Sichtung der theoretischen Befundlage erschien die visuelle Modalität am geeignetsten, um die Problemfelder der Sprachbedienung auszugleichen. Der zweite Optimierungsbereich soll durch die Einführung eines System Grounding Criterions adressiert werden. Dabei soll die Flexibilisierung des Dialogverhaltens anhand einer Bindung an die Erkennerkonfidenz geschehen, da diese einen robusten Schätzer der Systemsicherheit darstellt. Das Grounding-Element der Angleichung soll dagegen auf Wortebene umgesetzt werden, um Verständnisevidenzen auch verbal darzubieten.

Im folgenden Kapitel sollen aus diesen Lösungsvorschlägen die Kernfragestellungen der Arbeit abgeleitet und auf deren empirische Prüfung hingearbeitet werden.

5 Fragestellungen der Arbeit

In der bisherigen theoretischen Abhandlung konnte demonstriert werden, dass SDS eine vielversprechende Lösung für die Bedienung von fahrfremden Infotainmentfunktionen sein können (vgl. Kapitel 2). Dennoch bestehen eine Reihe technischer und gestalterischer Unzulänglichkeiten, die zu Einschränkungen in der Bedienbarkeit derzeitiger SDS führen und die nutzerseitige Akzeptanz schmälern.

Diese Arbeit adressiert daher die Steigerung der Gebrauchstauglichkeit automotiver SDS. Dabei werden die Verbesserungen nicht durch eine technisch-orientierte Weiterentwicklung, welche sich maßgeblich der Erhöhung der Spracherkennergenauigkeit und der Verständlichkeit der Sprachausgabe widmen würde, motiviert. Der hier präsentierte Vorschlag fokussiert darauf, ein nutzerfreundliches System durch adäquates Dialogdesign entstehen zu lassen. Insbesondere die Messgrößen Effektivität und Effizienz sollen allein durch das Dialogdesign Verbesserung finden und sich positiv auf die Nutzerzufriedenheit auswirken. Diese Zielsetzung hat die gestalterische Perspektive mit der technischen gemein.

Im Folgenden soll die Motivation zur Etablierung neuer Gestaltungsrichtlinien beim Design von SDS verdeutlicht werden, um anschließend allgemeine Fragestellungen für die empirische Untersuchungen ableiten zu können.

5.1 Motivation

Obwohl das Design eine bedeutende Rolle bei der nutzerseitigen Akzeptanz der Schnittstelle spielt, ist vergleichsweise wenig Grundlagenforschung zur **Dialoggestaltung** von nutzerfreundlichen Sprachdialogsystemen vorhanden. Auch aus den recht abstrakten Richtlinien der ISO Normen (Kap. 2.1.3) lässt sich nur begrenzt ableiten, wie SDS optimal gestaltet werden können. Zwar lassen sich durch diese Richtlinien Mängel im Bereich der Effizienz, Effektivität und Zufriedenheit identifizieren, aber konkrete Verbesserungsvorschläge für Dialoge können nicht zufriedenstellend extrahiert werden.

Hinzu kommt, dass aktuelle SDS im Fahrzeug nur eine ergänzende Bedienmodalität darstellen und sich so bisweilen häufig an den Gestaltungsrichtlinien von grafischen Nutzerschnittstellen orientieren. So finden sich viele Sprachdialogsysteme, die den Nutzer dazu zwingen, sich fest an Menüstrukturen zu halten, die der Logik und dem Design von grafisch-haptischen Bedienschnittstellen entsprechen.

Aufgrund des Mangels an adäquaten Gestaltungsrichtlinien für SDS wird in der vorliegenden Arbeit argumentiert, sich am **Vorbild** des nahezu perfekt eingeschwungenen zwischenmenschli-

chen Dialogs zu orientieren und Anregungen abzuleiten. Auch Dix et al. (2004) weisen darauf hin, dass der Erfolg von Systemen primär davon abhängt, ob Erwartungen und soziale Normen aus der menschlichen Kommunikation auf die Mensch-Maschine-Interaktion anwendbar sind. Da Dialoge zwischen Menschen nahezu ohne Kontrollaufwand geführt werden können, bieten sie eine optimale Ausgangsbasis, um nahezu belastungsfreie und nutzerfreundliche Dialoge zu gestalten. Vor dem Hintergrund, dass Nutzer intuitiv versuchen, die bekannten Dialogprinzipien in dem Mensch-Maschine-Dialog anzuwenden (vgl. Kapitel 2.3.4), erscheint eine Implementation dieser Strategien in ein SDS im Fahrzeug nicht nur sinnvoll, sondern notwendig.

Aufbauend auf diesen Annahmen wurde in Kapitel 3 eine Analyse zwischenmenschlicher Dialogführung angeschlossen. Als wesentlicher Erfolgsfaktor wurde dabei der kollaborative Charakter der Konversation identifiziert und der Prozess des Groundings beschrieben, in dem bestimmte Elemente ganz wesentlich zu effizienter und effektiver Kommunikation beitragen. Im Konkreten konnte aufgezeigt werden, dass in zwischenmenschlichen Dialogen umfangreiche Feedbackprozesse, eine situationsabhängige Anpassung des Dialogverhaltens und eine Angleichung an die Wortwahl des Gesprächspartners häufig dem Aufbau einer geteilten Wissensbasis dienen. In Kapitel 4.1 zeigte eine Analyse aktueller SDS im Hinblick auf die Einhaltung dieser Grounding-Prinzipien große Mängel. Die Sichtung der bestehenden Forschungsgrundlagen konnte darüber hinaus belegen, dass viele Fehler, die in Mensch-Maschine-Dialogen auftreten, als Folgen dieser mangelnden Kollaboration erklärt werden können (vgl. Brennan, 1998).

Der primäre Ansatz der Arbeit ist es nun zu prüfen, inwieweit eine Implementation der zentralen Grounding-Elemente (Feedback, System Grounding Criterion und Alignment) die Interaktion mit Sprachdialogsystemen hinsichtlich der Gebrauchstauglichkeit verbessern und damit deren Akzeptanz erhöhen kann. Kapitel 4.2 liefert bereits erste Hinweise darauf, dass eine Übertragung der zwischenmenschlichen Kommunikationsprinzipien vielversprechend erscheint.

Dabei wurde sich für die konkrete Umsetzung dieser drei Elemente der zwischenmenschlichen Dialogführung entschieden, da sie trotz adäquatem technischen Implementierungsaufwand durch ihre bedeutsame Rolle in zwischenmenschlichen Dialogen vielversprechend erscheinen, um den Mensch-Maschine-Dialog nutzerfreundlicher zu gestalten.

So wird davon ausgegangen, dass die Umsetzung von umfangreichen Feedbackprozessen und einer adaptiven Dialoggestaltung die Balance zwischen den Kommunikationspartnern herstellen, die Grounding-Prozess erleichtern und den kollaborativen Charakter der Kommunikation gerecht werden kann. Es ergeben sich drei globale Hauptannahmen, die die Gestaltung von SDS nach zwischenmenschlichen Kommunikationsstrategien betreffen und die empirischer Überprüfung bedürfen.

Einerseits wird postuliert, dass eine effektivere, effizientere und zufriedenstellendere Mensch-Maschine-Interaktion unabhängig von einem festgelegten Anwenderkreis entstehen kann, indem man die menschliche Verwendung von Sprache in Dialogen als Gestaltungsrichtlinie anwendet.

1. **Alle Facetten der Gebrauchstauglichkeit nach DIN EN ISO 9241-11 (1998) können durch die Erleichterung der Grounding-Prozesse verbessert werden.**

Eine Anlehnung an die zwischenmenschliche Kommunikation und damit die natürlichere Gestaltung von SDS führt zu geringerer Ablenkung, da die automatisierten Prozesse der Dialogführung nicht unterdrückt werden müssen. Ein ausgereiftes Dialogdesign nach menschlichem Vorbild kann demnach die mentalen Ressourcen des Fahrers schonen.

2. **Das mentale Beanspruchungslevel einer normalen zwischenmenschlichen Kommunikation wird nicht überschritten, wenn das Grounding zwischen Mensch und Maschine erleichtert wird.**

Postuliert wird weiterhin, dass (unabhängig von der Qualität der zugrundeliegenden Spracherkennungstechnologie) ein Design, welches die kollaborativen Aspekte von Sprache berücksichtigt, Fehlerkennungen und Fehlbedienungen vermeiden kann.

3. **Fehlerkennungen und spezifische Bedienfehler können durch die Umsetzung der Grounding-Elemente vermieden werden.**

Gerade die letzte Forschungsfrage verdeutlicht, dass hier die Usability-zentrierten und technisch-orientierten Entwicklungsstränge zusammenlaufen. Es soll in diesem Zusammenhang nachgewiesen werden, dass auch ein gutes Dialogdesign indirekt die Ziele der Effektivitätssteigerung adressieren kann.

Im folgenden Abschnitt werden spezifischere Forschungsfragen abgeleitet, die die Durchführung von drei empirischen Erhebungen motivieren.

5.2 Fragestellung der einzelnen Studien

Im Rahmen der ersten Untersuchung soll die Bedeutung des **Feedbacks**, also der Rückmeldungen von Empfängerzuständen im Mensch-Maschine-Dialog eruiert werden. Dabei wird sowohl die kombinierte, als auch die alleinige Rückmeldung von Systemzuständen und Dialoginhalten zu einem Referenzsystem verglichen, welches die Dialogspuren nach Clark (1996) nur unzureichend vermittelt.

Innerhalb einer Simulatorstudie soll gezeigt werden, dass die Interaktion mit einem SDS effizienter und angenehmer ist, wenn dieses mit einer ähnlichen Responsivität wie Menschen ausgestattet ist (vgl. Johnstone et al. 1995). Es wird davon ausgegangen, dass erfolgreiches Grounding zwischen Nutzer und Dialogsystem nur dann erfolgen kann, wenn sowohl Dialoginhalte als auch Zustandsfeedback vermittelt werden.

Die zweite Untersuchung adressierte hingegen die Implementation eines **System Grounding Criterions**. Dabei wird postuliert, dass durch diese Umsetzung des Prinzip des Least Collaborative Effort (Clark & Wilkes-Gibbs, 1986) der Grounding-Prozess erleichtert wird und sich die nutzerseitigen Aufwandskosten minimieren (Clark & Brennan, 1991). Im Rahmen einer Nutzerstudie soll gezeigt werden, dass ein System, das auf Basis der Erkennerkonfidenz dynamische Interaktionsstrukturen zeigt, die Gebrauchstauglichkeits-Kriterien nach DIN EN ISO 9241-11 (1998) besser erfüllt, als ein System mit statischem Rückfrageverhalten. Im Speziellen soll das in Kapitel 4.2.2 eingeführte Konsistenzdilemma untersucht werden und die allgemeine Frage nach der obligatorischen Systemkonsistenz für sprachliche Schnittstellen beantwortet werden.

In Studie III stehen die Effekte des systemseitigen **Alignments** im Vordergrund. Auch in dieser Studie wird davon ausgegangen, dass die Umsetzung des Grounding-Elements der lexikalischen Angleichung die nutzerseitige Akzeptanzbewertung steigern kann. Abgeleitet aus den in Kapitel 3 vorgestellten Untersuchungen des zwischenmenschlichen Dialogverhaltens soll geprüft werden, wie die funktionalen Ursachen des Alignments im Mensch-Maschine-Dialog gewichtet sind. Da bisherigen Studien (z. B. Porzel, 2006) auf die Untersuchung von langen Problemlösedialogen fokussierten, wurde der Effekt von systemseitigen Alignment in kurzen kommandoartigen Interaktionen, wie sie bei automotiver Sprachbedienung üblich sind, bisher noch nicht untersucht. Darüber hinaus ging die lexikalische Angleichung immer mit einer Effizienzverbesserung des Systems einher, die sich auch positiv auf die Nutzerbewertungen auswirken kann. Der alleinige Effekt einer lexikalischen Systemangleichung ohne Zeiteinsparungen wurde bisher ebenfalls nicht umfassend untersucht. Im Rahmen einer Probandenstudie soll die Forschungsfrage adressiert werden, ob linguistisches Alignment auch in kurzen Problemlösedialogen ohne objektive Effizienzverbesserung zu einer verstärkten nutzerseitigen Zufriedenheit und besseren Systembeurteilungen führt.

In dem folgenden Kapitel werden diese drei Studien dargestellt.

6 Empirischer Teil: Feedback, System Grounding Criterion und Alignment

Im Rahmen von drei experimentellen Nutzerstudien soll in diesem Kapitel gezeigt werden, dass Systeme, die auf Basis der Collaborative Theory zwischenmenschliche Kommunikationsstrategien anwenden, die Usability-Kriterien besser erfüllen als ein System ohne diese Anpassungen.

In den folgenden Studien wurde jedes der drei Umsetzungskonzepte separat evaluiert, um aufzuzeigen, welche zwischenmenschliche Kommunikationsstrategie das größte Potential zur Verbesserung der Gebrauchstauglichkeit bei einem Transfer auf ein SDS aufweist. Die Effektivitätsprüfung und die nutzerseitige Systembewertung erfolgten unter realitätsnahen Anwendungsbedingungen im Fahrsimulator. In einem abschließenden Kapitel werden die Ergebnisse der drei Hauptstudien miteinander in Beziehung gesetzt und Gestaltungsmaßnahmen für zukünftige automotive Sprachdialogsysteme abgeleitet.

Neben der Beschreibung der Operationalisierung der drei Gebrauchstauglichkeitsmaße Effektivität, Effizienz und Zufriedenheit in Form objektiver und subjektiver Variablen werden im Rahmen dieses Kapitels bedeutsame Effekte der Systemgestaltung auf diese Kriterien der Gebrauchstauglichkeit berichtet. Als Rahmenbedingung der Evaluation diente dabei stets die europäische Norm DIN EN ISO 9241-11 (1998). Der technisch orientierte Blick auf die tatsächliche Leistung des Systems wird dabei zu der empfundenen Leistung durch den Nutzer ergänzt. So galt es im Sinne der Effektivität zu eruieren, wie leicht bzw. schwer die Bedienung dem Nutzer fiel. Dazu wurden neben Fehlerkennungen und Fehlbedienungen auch subjektive Bewertungskriterien erhoben. Ebenso wurde die Effizienz über die Dialogdauer und deren Angemessenheitsbeurteilung erfasst. Nicht zuletzt bildet auch die Zufriedenheits- und Akzeptanzbeurteilung der Nutzer in jeder Untersuchung einen großen Bestandteil. Zur Erfassung jener subjektiven Bewertungen kamen Fragebögen, Interviews und indirekte Methoden (z. B. Rekognitionstests) zum Einsatz. Die Fragebögen bestanden dabei sowohl aus standardisierten Elementen als auch aus dynamisch zusammengestellten Abschnitten, die auf die spezifische Fragestellung der jeweiligen Untersuchung zugeschnitten waren. Nur wenige Einstellungen wurden bereits vor der Nutzung des Systems erfragt (Nutzungsbereitschaft, Erfahrung, Technikaffinität).

Im Folgenden soll neben einer allgemeinen Beschreibung des verwendeten Sprachdialogsystems die Implementation der Strategien beschrieben werden. Dabei werden die Umsetzungen aus der Collaborative Theory hergeleitet, Umsetzungsalternativen diskutiert und teilweise anhand von Vorstudien begründet. Durch die Ergebnisdokumentation und -diskussion der einzelnen Studien sollen die Anforderungen und Empfehlungen an die Gestaltung eines Sprachdialogsystems abgeleitet werden.

6.1 Studie I: Feedback

Kapitel 4.1.1 konnte aufzeigen, dass grafisches Feedback für Sprachdialogsysteme im Fahrzeug in aktuellen Systemen in unzureichender Form zur Darstellung von Systemzuständen genutzt wird. Der erste Abschnitt des empirischen Teils betrachtet daher im Rahmen eines nutzerorientierten Gestaltungsansatzes die Einführung umfangreicherer Zustandsmeldungen und die Anzeige von Dialoginhalten und motiviert damit neue Visualisierungskonzepte, die in einer Nutzerstudie hinsichtlich Machbarkeit, Usability und Ablenkungspotential untersucht werden. Das primäre Ziel der Visualisierungen ist es dabei, den Nutzer während und nach seiner Eingabe kontinuierlich Feedback zu geben. Dabei wird das Untersuchungsdesign der Studie I die Möglichkeit bieten, sowohl die kombinierte als auch die alleinige Anzeige des Inhalts- und Zustandsfeedbacks mit einem Referenzsystem ohne ergänzende Visualisierungen zu vergleichen. Es ergeben sich die folgenden Fragestellungen:

6.1.1 Hypothesen

Durch Empfangsanzeigen soll das System im Sinne der Collaborative Theory seinen Aufgaben als Empfänger besser nachkommen und das Grounding mit dem Nutzer erleichtern. In Anlehnung an Clarks (1996) zwei Spuren des Dialogs wird davon ausgegangen, dass erfolgreiches Grounding zwischen Nutzer und Dialogsystem nur dann erfolgen kann, wenn sowohl Inhalts- als auch Zustandsfeedback vermittelt werden. Durch die kombinierte Rückmeldung werden die Fähigkeiten und Grenzen des Systems zum einen und die zulässigen Interaktionsabläufe zum anderen transparent dargestellt. Darüber hinaus ermöglicht das System Reparaturen, indem das System nachfragt und anzeigt, wenn ein Input nicht interpretiert werden kann. Die Übernahme der kollaborativen Empfängerverantwortung soll die Einschätzung der Gesamt-Usability und Systemtransparenz vor allem dann erhöhen, wenn beide Feedbackarten kombiniert dargestellt werden.

Obwohl davon ausgegangen wird, dass alleiniges Zustands- und auch Inhaltsfeedback bereits zu einer besseren Bewertung der allgemeinen Gebrauchstauglichkeit des Systems führen, werden die besten Resultate somit bei einer kombinierten Darbietung erwartet:

1. **Ein System mit kombinierter, grafischer Darbietung von Inhalts- oder Zustandsinformationen erzielt bessere Bewertungen der allgemeinen Gebrauchstauglichkeit als ein System ohne Visualisierungen.**

Weiterhin wird davon ausgegangen, dass sowohl die Rückmeldung der Zustände des Systems als auch die Dialoginhalte zu einer besseren Bewertung der Systemtransparenz führen:

2. **Eine zusätzliche Rückmeldung erhöht die Systemtransparenz des Sprachdialogsystems.**

In bestimmten Aspekten der Gebrauchstauglichkeit und Systembeurteilung wird allerdings davon ausgegangen, dass die Vermittlung von Inhaltsfeedback andere Verbesserungen erzielt, als die Vermittlung von Zustandsfeedback.

So wird postuliert, dass durch die visuelle Darbietung der Systemstatus zulässige Handlungsoptionen aufzeigt und dadurch Fehlbedienungen reduziert werden. Die Anzahl von Nutzereingaben während das System spricht oder verarbeitet, sollten durch die transparente Bereitschaftsanzeige deutlich reduziert werden:

3. **Wenn Zustandsfeedback gegeben wird, reduziert sich die Anzahl der Fehlbedienungen (TTE-Fehler) und die Effektivität des Systems wird erhöht.**

Durch die Zustandsvisualisierung werden die zur Statusvermittlung notwendigen Merkmale kontinuierlich zurückgemeldet und damit die Bedingungen für ein erfolgreiches Grounding optimiert (Clark & Brennan, 1991). Die Gesprächsinitiierungskosten nehmen ab, da durch eine Animation des Ready-Zustandes deutlicher wird, wann das System empfangsbereit ist. Die Wartekosten während des Verarbeitungszustands werden durch die Vermittlung des Status zwar nicht verkürzt, reduzieren aber die Unsicherheit des Nutzers und damit die Wahrscheinlichkeit einer Fehlbedienung (damit auch mögliche Fehlerkosten). Diese Reduktion der nutzerseitigen Aufwandskosten und die Unterstützung des Grounding-Prozesses sollten zu einer höheren nutzerseitigen Zufriedenheit führen:

4. **Eine zusätzliche Visualisierung der Systemzustände erhöht die nutzerseitige Zufriedenheit.**

Es wird angenommen, dass das Inhaltsfeedback durch die positive Verständnisevidenz die Systembeurteilung nicht nur hinsichtlich der Transparenz verbessert. Durch negative Evidenzen, gegeben in dem Fall es wird etwas angezeigt, was der Nutzer nicht gesagt hat, bietet es ihm die Gelegenheit zu einer frühen Korrektur der Eingaben (Effizienz). Dies sollte sich positiv auf die Effizienz auswirken, da die Reparaturkosten des Dialogs somit reduziert werden können (siehe Clark & Brennan, 1991):

5. **Eine zusätzliche textuelle Visualisierung der Dialoginhalte erhöht die Effizienz.**

Durch die Persistenz der Anzeige und der Möglichkeit die Nachricht später noch einmal anzusehen, werden die Bedingung zur Etablierung eines CG verbessert (Kap. 3.2.1; ebd.). Diese Vermeidung der Vergänglichkeit kann die mentale Beanspruchung, die aufgrund umfangreicher

sprachlicher Ausgaben auftreten kann, reduzieren, da der Nutzer selbst entscheiden kann, wann er seine Aufmerksamkeitsressourcen zur Systeminteraktion verwendet:

6. **Eine zusätzliche textuelle Visualisierung der Dialoginhalte verringert die mentale Belastung der sprachlichen Interaktion.**

Da das Inhaltsfeedback den Common Ground textuell darbietet, wird davon ausgegangen, dass der Nutzer ein besseres Verständnis der Systemkompetenzen und -grenzen entwickeln und mehr Kommandos memorieren können:

7. **Eine zusätzliche (textuelle) Visualisierung der Dialoginhalte erhöht die Anzahl der erinnerten Kommandos.**

Die grafische Umsetzung beider Visualisierungsarten spricht vorherige Erfahrungen und Schemata der Nutzer an und ist gleichermaßen verständlich für das Zustands- und das Inhaltsfeedback:

8. **Es zeigen sich keine Unterschiede in der Bewertung der Gestaltung des Inhalts- und Zustandsfeedbacks.**

Die kontrastreiche Visualisierung und gute Platzierung des visuellen Feedbacks sollen dem Nutzer helfen, den Systemzustand wahrzunehmen und zu interpretieren. Da die farbigen Zustandsanzeigen hinsichtlich verschiedener Wahrnehmungsaspekte optimiert wurden, kann auch davon ausgegangen werden, dass eine periphere Wahrnehmung nahezu ohne Blickabwendung möglich ist. Aufgrund des Leseaufwands ist nur für das Inhaltsfeedback mit einer erhöhten Blickabwendung zu rechnen:

9. **Die prozentuale Blickabwendung von der Fahrsituation erhöht sich durch die Visualisierungen nur bei der textuellen Repräsentation der Dialoginhalte.**

Die prozentuale Blickabwendung bei der textuellen Darbietung der Dialoginhalte kann jedoch durch eine adäquate Gestaltung kompensiert werden. In Anlehnung an Salmen (2002) wird davon ausgegangen, dass Blickabwendungen nutzerseitig nur kontrolliert vorgenommen werden und durch antizipatorische Wahrnehmung kompensiert werden können. Auch wenn die Nutzereingabe und die Systemausgabe vollständig angezeigt werden, vergleichbar zu der Umsetzung von Curin et al. (2011; Kapitel 4.2.1), wird nicht davon ausgegangen, dass es zu negativen Einflüssen auf die Fahraufgabe kommt:

10. **Weder Zustands- noch Inhaltsfeedback führen zu einer Beeinträchtigung der Fahraufgabe.**

Explorativ soll der Frage nachgegangen werden, ob eine der beiden Spuren für den erfolgreichen und zufriedenstellenden Dialogablauf innerhalb der Mensch-Maschine-Interaktion eine größere Rolle spielt.

In dem folgenden Kapitel soll der Untersuchungskontext, der zur Überprüfung der Fragestellungen Anwendung fand, näher erläutert werden.

6.1.2 Methode

Zunächst werden die Fahr- und Sprachaufgaben beschrieben. Nach einer detaillierten Erläuterung der Umsetzung des Referenz-SDS, welches in allen Studien Anwendung fand und der Feedbackarten, werden der Versuchsplan und die Erhebungsinstrumente eingeführt und begründet. Anschließend wird das Untersuchungsdesign, die Stichprobenzusammensetzung, der Versuchsablauf und die Rahmenbedingungen der Auswertung erläutert.

6.1.2.1 Fahraufgabe

Die Untersuchung wurde während einer Folgefahrt im feststehenden **Fahrsimulator** der Volkswagen Konzernforschung durchgeführt (siehe Abbildung 17). Dieser besteht aus einer rudimentären Nachbildung eines Fahrzeugs mit Automatikgetriebe und Fahrersitz, Lenkrad, Pedalerie und Ganghebel, einer Windschutzscheibe mit Seiten- und Rückspiegeln, sowie einem freiprogrammierbaren Kombiinstrument mit Geschwindigkeitsanzeige und Drehzahlmesser. Diese „Sitzkiste“ ist statisch, das heißt, in der Längs- und Querführung unbeweglich, so dass Fahrbewegungen nur optisch simuliert werden. Die Streckensimulation wird auf drei Leinwänden dargestellt, die vor der Sitzkiste im Halbkreis aufgebaut sind. Das Fahrersichtfeld beträgt hierbei ungefähr 120 Grad.

Abbildung 17: Versuchsaufbau im Volkswagen Fahrsimulator

Als Fahrtstrecke wurde ein Landstraßen-Szenario von 37 Kilometern Länge gewählt, mit geringem Verkehrsaufkommen (25%), leichten Kurven und zwei Ortsdurchfahrten von insgesamt 1.5 km Länge. Die geringe Streckenschwierigkeit wurde gewählt, um eine möglichst realistische Bediensituation mit hoher Nutzungsbereitschaft für das System seitens der Fahrer zu simulieren. Die getesteten Versuchspersonen wurden instruiert, einem Vorderfahrzeug mit möglichst kon-

stanter Geschwindigkeit von 70 km/h und einem Abstand von 50 Metern zu folgen. In den Aufgabenfahrten begannen die Aufgabeninstruktionen erst nachdem die Probanden zum Vorderfahrzeug aufgeschlossen hatten.

Während der Fahrten wurden **Fahr- und Blickdaten** mit Hilfe des Dikablis-Systems aufgezeichnet (Ergoneers, 2011). Hierbei handelt es sich um ein Hard- und Softwaresystem der Ergoneers GmbH. Es ermöglicht die störungsarme Aufzeichnung von Blickdaten mit Hilfe einer am Kopf der Probanden angebrachten Brille (engl. *Head-Mounted-System*). Die Blickdaten stellten ein objektives Maß der Ablenkung dar, welches insbesondere bei der Hinzunahme von Visualisierungen von Bedeutung ist.

Bei der Aufzeichnung der Fahrdaten wurde sich in allen Studien primär auf die Spurhaltung konzentriert. Die Spurhaltung wurde gewählt, da sie im Vergleich zu anderen Fahrparametern ein Maß darstellt, welches sensibel auf verschiedene Arten von Ablenkung reagiert (Zwahlen, Adams, DeBald, 1988). Als Maß wurde sich in den Untersuchungen auf die Standardabweichung der lateralen Position (SDLP) bezogen. Sie berechnete sich aus der Standardabweichung aller gemessenen Abstände des Fahrzeugmittelpunkts zum rechten oder linken Fahrbahnrand. Je größer die Werte ausfallen, umso stärker schwankt ein Fahrer um die mittlere laterale Position in der Spur (Knappe, Keinath & Meinecke, 2006).

Zunächst soll das Referenzsprachdialogsystem vorgestellt werden, welches als Basis für alle im Folgenden beschriebenen Umsetzungen diente.

6.1.2.2 Sprachdialogsystem

Bei dem Basissystem handelte es sich um ein Sprachdialogsystem, dem ein automatischer Spracherkenner auf Basis von Hidden–Markov–Modellen zugrunde lag (siehe Kap 2.2.1). Die Syntax einzelner Phrasen oder ganzer Sätze wurde durch ein statistisches Sprachmodell (SLM) abgebildet. Das Dialogsystem wurde in allen Bedingungen durch das Drücken des Push-to-Talk-Knopfs am Lenkrad aktiviert.

Das verwendete Dialogsystem verfolgte einen sogenannten **formularbasierten Ansatz** (engl. *form based*). Die Formulare (engl. *forms*), auf deren Basis das Dialogsystem modelliert ist, beschreiben die Funktionalitäten des Hintergrundsystems. Die Ausführung eines Formulars ist also mit der Ausführung einer Funktion (z. B. einen Telefonanruf) gleichzusetzen. Die zu Beginn der Interaktion noch leeren Felder (engl. *fields*) einer Formularvorlage müssen dabei durch den Nutzer gefüllt werden, um eine Funktion ausführen zu können. Die einzelnen Informationen, die nutzerseitig eingegeben werden, füllen die Felder, die wiederum immer einem bestimmten Formular zugewiesen sind. Die Reihenfolge der Dialogschritte ist dabei systemseitig nicht festgelegt und

kann je nach Vollständigkeit der Nutzereingabe variieren. Das SDS zeigt somit eine gemischte Initiative (vgl. Kap. 2.2.4) und erlaubt dem Nutzer eine natürlich-sprachliche und freie Äußerung von Eingaben.

Bei der anfänglichen Eingabe bestimmt der Nutzer durch seine Äußerung ein bestimmtes Formular und kann bereits einige Felder füllen. Bei der Eingabe *Helga Otto zu Hause anrufen.* würde es sich bei *anrufen* um die auszuführende Funktion, also um das Formular handeln. Mit den Feldern *Helga Otto* (Name) und *zu Hause* (Nummernkategorie) füllt der Nutzer bereits alle notwendigen Felder zur Ausführung des Formular. Würde der Nutzer nur *Helga Otto anrufen.* sagen, so würde das System nach dem fehlenden Feld, der Nummernkategorie fragen, sofern im Adressbuch mehrere Nummern von diesem Kontakt hinterlegt sind.

Somit kann sowohl der Nutzer den Dialog kontrollieren, indem er proaktiv eine Mehrzahl an Feldern füllt, als auch das System, das die Initiative ergreift, wenn notwendige Informationen zur Ausführung der Aktion fehlen oder widersprüchlich sind. Für jedes Feld können spezifische Prompts definiert werden, um die fehlenden Informationen zu erfragen oder kontextspezifische Hilfe anzubieten.

Das System ordnet die erkannten Details dabei automatisch den erforderlichen Informationen zu. Dem Dialogmodell liegen immer Informationen darüber vor, in welchen Kontext sich der Dialog gerade befindet und welche Felder bereits gefüllt sind. Darüber hinaus prüft das System auch die Konsistenz der Eingabe. So wird es beispielsweise alle verfügbaren Nummernkategorien von Helga Otto vorlesen, wenn die Kategorie *zu Hause* nicht im Adressbuch hinterlegt ist. Innerhalb des Dialogs ist ein Überschreiben der Felder oder ein Formularwechsel ohne Unterbrechung möglich, wie das folgende Dialogbeispiel belegt:

Nutzer	Helga Otto zu Hause anrufen	Multi-Slot Eingaben
System	Nummernkategorie "zu Hause" ist für Helga Otto nicht gespeichert. Möchten Sie Helga Otto mobil anrufen?	Konsistenzprüfung
Nutzer	Nein, SMS schreiben.	Formularwechsel
System	Bitte diktieren Sie die Nachricht.	Formularausführung

Diese Art des gemischt-initiierten formularbasierten Dialogmodells bietet eine hohe Flexibilität im Dialogablauf. Funktionsseitig beschränkt sich die Umsetzung allerdings auf die Bedienung des **Telefonbuchs**. So können per Sprache Telefonanrufe gestartet, Nachrichten diktiert oder vorgelesen, Nummern eingegeben oder geändert und die Navigation zu Kontakten angesteuert werden. Aufgrund der hohen Varianz von Telefonbucheinträgen und dem limitierten Funktionsumfang

wurden im vorliegenden System ausschließlich TTS Prompts generiert und keine voraufgezeichneten Sprachausgaben verwendet.

Die Sprachausgaben des Systems bedienten sich in seiner ursprünglichen Variante immer gleicher Wortwahl. Der Dialog erfolgte über die sprachliche Modalität und zeigte keine (visuellen oder akustischen) Empfangsbestätigungen. Abseits einer Tonfolge wurden keine dialogstrukturierenden Signale gesendet, um den Aktivitätszustand des Systems anzuzeigen. Es verfolgte eine defensive Interpretationsstrategie der Erkennerergebnisse. Der Dialogablauf wurde stets durch eine finale Bestätigungsanfrage an den Nutzer abgeschlossen. Stark zusammengefasst zeichnet sich das Referenzsystem demnach durch die folgenden Eigenschaften aus und bildet damit typische Umsetzungen von automotiven SDS ab:

- Keine Rückmeldungen des Empfängerstatus
- Defensive Interpretationsstrategie (mindestens eine finale Bestätigungsanfrage)
- Gleichbleibende Wortwahl der Systemprompts

Wie bereits einleitend erwähnt, wurde jeweils eines dieser Merkmale in den folgenden Studien adressiert, durch zwischenmenschliche Kommunikationsstrategien ersetzt und zur ursprünglichen Konfiguration verglichen. Im Fokus der ersten Studie steht die Umsetzung von umfangreichen Rückmeldeprozessen.

6.1.2.2.1 Umsetzung Feedback

In Kapitel 3.2.1 wurde deutlich, dass es für eine effektive Dialogführung wichtig ist, die Sprecher über den Zustand ihres Gesprächspartners zu informieren. Ein menschlicher Empfänger gibt Feedback während er dem Sprecher zuhört oder unmittelbar während seines Übergangs von der Hörer- in die Sprecherrolle (McTear, 2004, S. 54). Er signalisiert dabei einerseits, dass er zuhört (Zustandsfeedback), andererseits, was er verstanden hat (Inhaltsfeedback). In Anlehnung an Clarks (1996) zwei Arten von Dialogspuren (siehe Kapitel 3.2.1) soll nicht nur jede Phase der Sprachproduktion durch das System vermittelt, sondern auch die korrespondierenden Empfangshandlungen als Akzeptanzevidenzen präsentiert werden. Beide Dialogspuren sollen folglich in einem SDS Umsetzung finden.

Das Prozess- oder **Zustandsfeedback** (ZF) soll den Nutzer darüber informieren, was das System gerade tut (prozessbegleitende Aspekte) bzw. welche Dialogaktivitäten gerade erwartet werden (dialogstrukturierende Aspekte). Innerhalb eines Dialogschrittes lassen sich fünf Systemzustände identifizieren, die dem Nutzer vermittelt werden sollen (siehe Abbildung 18).

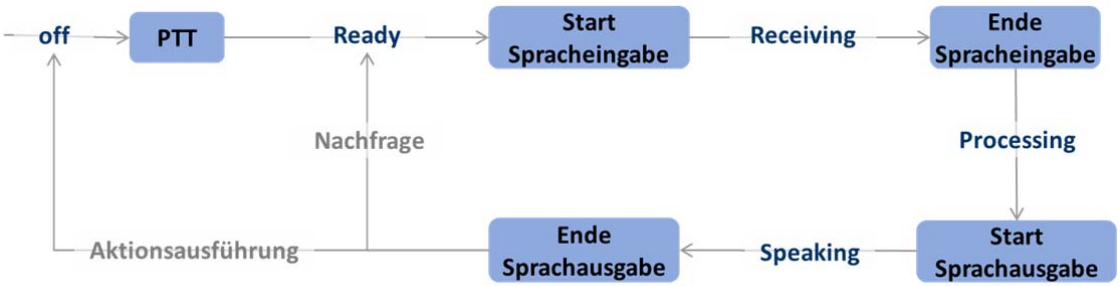

Abbildung 18: Systemzustände innerhalb eines Dialogs

Die Zustände beziehen sich vor allem auf die verschiedenen Aktivitätszustände des SDS und können sowohl als dialogstrukturierende, als auch als prozessbegleitende Feedbacksignale fungieren (siehe Tabelle 9).

Tabelle 9: Beschreibung der Systemzustände

	Zustand	**Beschreibung**	**Sprech-Verantwortung**	**Signalart**
0	Off	Erkenner geschlossen		Dialogstrukturierend
1	Ready	Erkenner offen	Nutzer	Dialogstrukturierend
2	Receiving	Empfang des Sprachsignals	Nutzer	Prozessbegleitend
3	Processing	Verarbeitung des Sprachsignals	System	Prozessbegleitend und dialogstrukturierend
4	Speaking	Sprachausgabe des Systems	System	Prozessbegleitend und dialogstrukturierend

Damit werden die Phasen des SDS zurückgemeldet, in denen es Eingaben erfassen kann, Eingaben verarbeitet oder selbst spricht. Vor allem die Zustände Receiving und Processing lassen sich als prozessbegleitende Feedbacksignale identifizieren (vgl. Kapitel 3.2.1). Es leiten sich auch zwei dialogstrukturierende Signale ab, die einen Sprecherwechsel initiieren. So ist der Nutzer nach dem Übergang des Systems vom deaktivierten (Zustand 0) zum aktivierten Zustand (Zustand 1) angehalten seine Spracheingabe zu tätigen. Nach Abschluss der Eingabe hält das System den Turn, bis es durch eine Nachfrage die Verantwortung wieder dem Nutzer zurückspielt oder den Dialog durch eine Aktionsausführung abschließt. Der Ready-Zustand übergibt den Turn somit an den Nutzer, während der Processing-Zustand ihn aktiv beim System hält. Das folgende Diagramm (siehe Abbildung 19) stellt die möglichen Zustandsübergänge dar.

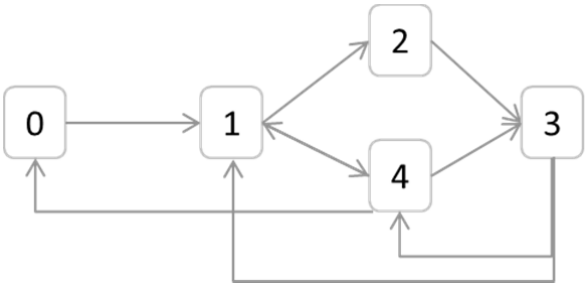

Abbildung 19: Zustandstransitionen

Verglichen zu bisherigen SDS, sollte das Zustandsfeedback mehr Informationen als eine bloße Bereitschaftsanzeige bieten und gerade im Rahmen von Verarbeitungszuständen der Vermeidung von Abbrüchen und unerwünschtem Input dienen.

In dem folgenden Abschnitt soll die **Umsetzung** der fünf Systemzustände des SDS erläutert werden. Die Visualisierung des Systemzustands erfolgte dabei in abstrakter Weise über einen Farbcode. Ein Vorteil, den diese rudimentäre Form der Anzeige bietet, ist die gute periphere Wahrnehmbarkeit eines Farbwechsels bei angemessener Farbwahl. Neben einem geringen Ablenkungspotential bietet die Anzeige die Chance durch einen geringen Umsetzungs- und Anzeigeaufwand alle fünf Zustände zu vermitteln. Die Größe der Umsetzung kann dabei an die realistischen Abmessungen, die eine Visualisierung für ein SDS in einem Kombidisplay beanspruchen dürfte, angelehnt werden. Dass Nutzer trotz der hohen Transferleistung ein mentales Modell der Systemzustände ausbilden und sie in Verbindung mit dem Sprachdialogsystem bringen können, konnte eine **Vorstudie** belegen (Anhang 12.1).

Aus jener Vorstudie ließen sich folgende basale Gestaltungsrichtlinien der abstrakten Systemzustandsvisualisierung ableiten:

- Eine Anzeige der verbleibenden Zeit sollte bei der Vermittlung des Verarbeitungszustandes (Processing) erfolgen. Bei prozessbegleitenden Zuständen (Receiving, Speaking) spielt sie eine untergeordnete Rolle.
- Warnfarben (insbesondere Rot) werden nicht als Aktivitätsindikator des Systems akzeptiert. Nutzer wenden ein mentales Modell, vergleichbar der Aufnahmefunktion es Videorekorders bei der Interaktion mit einem Spracherkenner nicht an.
- Das Zustandsfeedback sollte um Verständnisevidenzen ergänzt werden.
- Eine einfarbige Vermittlung der Systemzustände ist ausreichend. Animationen von Farbverläufen zur Verdeutlichung des Turn-Takes zwischen System und Nutzer sind nicht nötig.

Die Ergebnisse der Vorstudie sollten in einem erweiterten Rückmeldekonzept konsolidiert werden. Bei dem **Farbkonzept** wurde sich an den physiologischen Grundlagen des Farbsehens orientiert, um den Wahrnehmungsprozess optimal zu unterstützen (Herczeg, 1994). Für die Zustandsrückmeldung wurde neben der kontrastreichen Farbe Weiß auch die Farbe Blau gewählt, da diese bei den gegebenen Rahmenbedingungen viele wahrnehmungspsychologische Vorteile vereint. So ist sie bei guter Farbsättigung, ausreichender Helligkeit und gutem Kontrast in der Peripherie besonders gut erkennbar (Salmen, 2002). Die Zustände Ready und Receiving, innerhalb derer die Sprechverantwortung beim Nutzer lag, wurden in Weiß dargestellt. Wohingegen die Zustände Processing und Speaking, bei denen das System die Turnverantwortung trägt in einem hellen Blau dargestellt wurden.

Da insbesondere den Turn-Grenzen eine zentrale Bedeutung zugewiesen wurde, wird in der Feedback-Studie der Zeitraum nach dem Drücken des PPT-Knopfes bis zur Erkenneröffnung durch eine Animation aufgewertet, die ergänzend zu dem Jingle dargeboten wird. Der Aufbau einer weißen Punktereihe überbrückt die kurze Wartephase bis der Systemzustand Ready eintritt und dem Nutzer die Sprecherrolle zukommt (siehe Abbildung 20).

Abbildung 20: Ready-Animation

Anschließend wird der Ready-Zustand als weiße Punktereihe dargestellt. Tätigt der Nutzer dann seine Eingabe, so ändert sich der Zustand vom passiven Zuhören hin zum Empfang eines Sprachsignals (Receiving). Dabei unterschied sich der Receiving-Zustand nur durch ein sanftes Pulsieren der weißen Punktereihe von dem Ready-Zustand (siehe Abbildung 21).

Abbildung 21: Systemzustände Ready und Receiving

Nach Beendigung der nutzerseitigen Spracheingabe geht das System in den Verarbeitungszustand über (Processing). Dieser Zustand wurde als laufender blauer Punkt innerhalb der weißen

Punktereihe dargestellt (siehe Abbildung 22). Um die Darbietung des Processing-Zustandes konstant zu ermöglichen, wurde der Verarbeitungszustand künstlich auf eine Sekunde verlängert.

Abbildung 22: Systemzustand Processing

Der Verarbeitungsphase folgt die Antwort des Systems (Speaking). Auch jene wurde visuell unterlegt und als pulsierende Punktereihe dargestellt. Die farbliche Codierung war dabei von dem Empfängerzustand des Systems abhängig und stellt damit eine Erweiterung des Zustandsfeedback um Verständnisevidenzen dar. So erschien die Sprachausgabe durch eine blaue pulsierende Punktereihe, wenn das System davon ausging, die Eingabe verstanden zu haben. Im Falle einer Fehlerkennung verfärbte sich die Systemanzeige Orange[20] (siehe Abbildung 23). Die Differenzierung zu der, unter normalen Umständen hellblauen, Punktereihe geschieht somit über eine Farbe, die im orange-roten Bereich liegt und damit Aufmerksamkeit erzeugt.

Abbildung 23: Systemzustände Speaking (verstanden) und Speaking (nicht verstanden)

Neben dem Zustandsfeedback soll auch die erste Dialogspur eine visuelle Umsetzung finden. Die Rückmeldung der erkannten Inhalte soll dem Nutzer Sicherheit darüber vermitteln, dass das System die Spracheingabe wie beabsichtigt verstanden hat. Die Umsetzung des **Inhaltsfeedbacks** (IF) ist im Folgenden beschrieben. Zur Vermittlung der ersten Spur des Dialogs wurde eine Anzeige der vollständigen Nutzerein- und Systemausgabe innerhalb des Kombi-Displays verwendet.

[20] Auf die Warnfarbe Rot wurde im Kontext des Kombidisplays bewusst verzichtet, da diese auf einen schweren (technischen) Systemfehler hinweisen könnte oder direkt mit sicherheitskritischen Fahrerassistenzfunktionen verwechselt werden könnte.

Sobald der Nutzer seine Eingabe fertig eingesprochen hatte, wurde diese als weiße Sprechblase zentral im Kombiinstrument angezeigt (siehe Abbildung 24). Der Text der Sprechblase entsprach dabei den Inhalten, die das System verstanden hatte.

Abbildung 24: Inhaltsfeedback Nutzereingabe

Die Antwort des Systems erschien unterhalb der Nutzersprechblase in blau. Im Laufe der Interaktion wurden die Sprechblasen der vorangegangenen Interaktion transparenter und verschwanden aus der Anzeige, sobald eine neue Ein- oder Ausgabe vorgenommen wurde (siehe Abbildung 25). Somit wurde eine Gesamtanzahl von zwei Sprechblasen nie überschritten.

Abbildung 25: Inhaltsfeedback Systemausgabe (verstanden)

Konnte das System eine Eingabe nicht verstehen, so färbte sich die Systemsprechblase äquivalent zu der Zustandsanzeige orange (siehe Abbildung 26).

Abbildung 26: Inhaltsfeedback Systemausgabe (nicht verstanden)

Es wurde angestrebt, dass die visuellen Rückmeldungen des Systems der Nutzereingabe schnell und klar wahrnehmbar folgen. Allerdings konnte eine simultane Ausgabe technisch nicht umgesetzt werden, da der vorliegende Dialogmanager die Spracheingabe erst im Ganzen verarbeitet

haben musste, um sie dann als Textstring anzeigen zu können. Ähnlich verhielt es sich mit der Sprachausgabe des Systems. Diese konnte auch erst nach Abschluss des Prompts angezeigt werden. Für das Inhaltsfeedback konnte somit nur eine eingeschränkte simultan-redundante multimodale Interaktion erreicht werden (vgl. Kapitel 2.1.1). Die gleichen Informationen konnten nicht parallel, sondern nur konsekutiv angezeigt werden. Salmen (2002) konnte in diesem Zusammenhang allerdings nachweisen, dass eine solche konsekutive Ausgabe ausreichend ist, da die akustische und visuelle Präsentation lediglich im Wechsel wahrgenommen werden können.

Insgesamt wurde versucht die visuellen Anzeigen von Zustands- und Inhaltsfeedback so beeinträchtigungsfrei wie möglich umzusetzen. Bei der Gestaltung der textuellen Bestandteile wurde sich an der DIN EN 894-2 (1997), die Anforderungen an die Gestaltung von Anzeigen definiert, und Salmen (2002) orientiert, um die problemlose Rezeption zu unterstützen und das Ablenkungspotential so gering wie möglich zu halten. Ihr Ablenkungspotential soll in der folgenden Nutzerstudie jedoch spezifischer adressiert werden.

Kombiniert man das Zustands- und Inhaltsfeedback so werden sechs der acht Zustände des adaptierten Contribution Modell (Brennan & Hulteen, 1995, Kap. 4.1.1) angegeben (siehe Tabelle 10). Dabei werden nicht nur negative, sondern auch positive Verständnisevidenzen vermittelt. Die beiden letzten Zustände werden aufgrund der Untersuchungssituation nicht dargeboten. So werden die Kommandos, die zumeist den Telefonkontext betreffen, nicht bis zu einer realen Handlung ausgeführt. Zustand 6, der durch Wähltöne angezeigt werden könnte, wurde nicht erreicht, da der Dialog nach der Ankündigung der Handlungsabsicht durch das System als beendet galt.

Tabelle 10: Einordnung der Rückmeldungen in das adaptierte Contribution Model

Zustand	**Beschreibung**	**Darbietung**
Zustand 0	Das System ist nicht bereit.	Keine Punktereihe sichtbar.
Zustand 1	Das System ist bereit.	Aufbau Punktereihe.
Zustand 2	Das System empfängt.	Auf- und Abdimmen der Punktereihe.
Zustand 3	Das System erkennt.	Leere Sprechblase erscheint.
Zustand 4	Das System interpretiert.	Verstandener Text angezeigt.
Zustand 5	Das System hat eine Handlungsabsicht.	Gesprochener Text angezeigt.
Zustand 6	Das System führt aus.	x
Zustand 7	Das System berichtet das Ergebnis.	x

Die unterschiedlichen visuellen Rückmeldungen erfolgten über ein 12,3" Sharp Kombidisplay mit einer Auflösung von 1280 x 480. In diesem wurden, wie in Abbildung 27 ersichtlich auch die Geschwindigkeits- und Drehzahlmesseranzeige integriert.

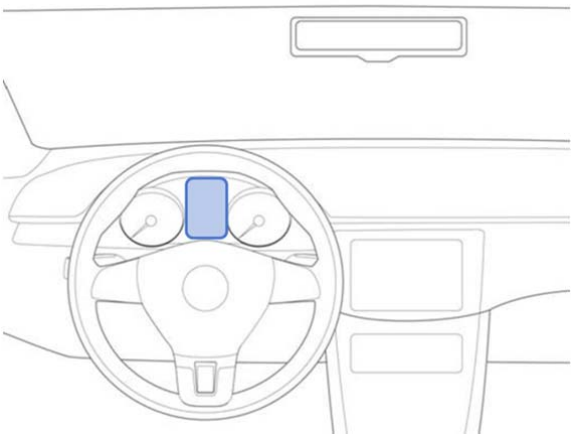

Abbildung 27: Anzeigeort

6.1.2.2.2 Kommandos

Bezüglich der **Sprachaufgaben** erhielten die Probanden vom Versuchsleiter (VL) verbal Aufgaben, die sie mit dem SDS während der Folgefahrt durchführen sollten. Der VL saß in einem Nebenzimmer, welches durch eine Scheibe Sichtkontakt sowie über Mikrofon und Lautsprecher Sprachkontakt zwischen Versuchsleitern und Probanden erlaubte (siehe Abbildung 28).

Abbildung 28: Versuchsleiterraum

Bei der Auswahl der acht Kommandos sollten drei Aspekte Beachtung finden. Einerseits wurde versucht möglichst viele **Funktionen** des Adressbuchkontextes wie Anrufen, Telefonkonferenz, Navigation zu einem Kontakt und SMS schreiben einzubeziehen.

Weiterhin waren **ambigue Namen** Bestandteil von drei Kommandos um etwas längere Dialoge zu provozieren. Dabei trat eine Systemnachfrage mit einer künstlichen Pause von einer Sekunde

auf. Durch diese prosodische Unterbrechung sollte ein Sprecherwechsel provoziert werden, der zu einer zu frühen Spracheingabe führt.

Ein beispielhafter Dialogablauf ist im Folgenden dargestellt:

Nutzer	Zu Linda navigieren.
System	Meinten Sie Linda Huber oder Linda Albrecht? <Pause> Bitte nennen Sie einen Nachnamen.
Nutzer	Huber.
System	Möchten Sie zu Linda Huber navigieren?
Nutzer	Ja.
System	Navigation wird gestartet.

Drittens sollte der Effekt des Vorsprechens der Kommandos durch den Versuchsleiter minimiert und eine freiere Wortwahl unterstützt werden, in dem die Aufforderung zur Spracheingabe bei den letzten vier Interaktionen in Form von **Szenarien** erfolgte. Zwei der vier Szenarien enthielten dabei so offene Aufforderungen, dass den Versuchspersonen kein Aktionswort (Anrufen oder Nachricht senden) vorgegeben wurde und sie selbst die Formulierung konstruieren mussten.

Insgesamt wurden den Probanden acht Kommandos und vier Szenarien vorgegeben. Damit ergaben sich 12 Interaktionen pro Versuchsperson. Die Kommandos und Szenarien unterschieden sich hinsichtlich der Anzahl notwendiger Sprecherwechsel. Die folgende Tabelle 11 verdeutlicht diese Differenz.

Tabelle 11: Kommandoübersicht Studie I

	Vorgabe	**Mindestanzahl TurnTakes**
Kommando 1	Sie möchten Ulrich Dietrich im Büro anrufen.	2
Kommando 2	Sie möchten ihre verpassten Nachrichten angezeigt bekommen.	1
Kommando 3	Sie möchten zu Laura Albrecht navigieren.	2
Kommando 4	Sie möchten Stefan auf dem Mobiltelefon erreichen.	3
Kommando 5	Sie wollen eine Telefonkonferenz mit Jürgen und Marta beginnen.	2
Kommando 6	Sie möchten eine SMS an Max schreiben	2
Kommando 7	Sie wollen sich zu Linda navigieren lassen.	3
Kommando 8	Sie möchten Helga Otto mobil anrufen	2

	Vorgabe	**Mindestanzahl TurnTakes**
Szenario 1	Sie sind mit ihrer Freundin Laura verabredet. Leider verspäten Sie sich, da sie noch etwas abholen müssen. Informieren Sie Laura, dass sie später kommen.	2
Szenario 2	Sie müssen ein Paket bei Hertha Otto abholen, wissen allerdings nicht, wie sie dort hinkommen. Nutzen Sie die Navigation.	2
Szenario 3	Ihre Freundin Laura wartet vor dem Kino auf Sie, aber Sie haben sich verfahren. Nun wollen Sie Laura nach dem Weg zu fragen.	2
Szenario 4	Laura kann es schlecht erklären. Lassen Sie sich zu Stefan navigieren.	3

Während bei dem Kommando *verpasste Nachrichten anzeigen* keine Bestätigungsanfrage durch das System vorgesehen und somit nur ein Sprecherwechsel nötig war, gab es Kommandos bei denen durch die Ambiguität der Namen (Stefan, Max, Linda) mindestens drei Turns notwendig waren um die Aktion auszuführen.

Um die Effekt des Inhalts- mit denen des Zustandsfeedbacks vergleichbar zu halten, wurden bewusst keine Kontexte verwendet, in denen die textuelle Darbietung von vornherein Vorteile aufweist. So wurde das Kommando 6 als erfolgreich beendet betrachtet, bevor es zu dem Diktieren einer Nachricht kam.

6.1.2.3 Versuchsplan

Da Lern- und Reihenfolgeeffekte zu erwarten waren, wurde für die Datenerhebung ein Zwischensubjekt-Design gewählt. Es erfolgte eine randomisierte Zuordnung der Probanden zu einer der vier Bedingungen des 2 x 2 Faktoren-Designs (Inhaltsfeedback: nicht präsent oder präsent x Zustandsfeedback: nicht präsent oder präsent). Neben einer Kontrollbedingung, in der keinerlei visuelle Rückmeldungen angezeigt wurden, gab es eine Bedingung, in der nur Inhaltsfeedback, eine in der nur Zustandsinformationen und eine Bedingung in der eine kombinierte Anzeige beider Feedbackarten dargeboten wurde. Der Versuchsplan ist in Abbildung 29 dargestellt. Innerhalb der jeweiligen Feedback-Bedingung führten alle Personen die gleichen Aufgaben in fester Reihenfolge durch.

Zur Beurteilung der allgemeinen Effekte der Sprachbedienung auf die Fahr- und Blickparameter fanden zwei Referenzfahrten ohne Sprachaufgaben jeweils vor und nach dem Aufgabenteil statt.

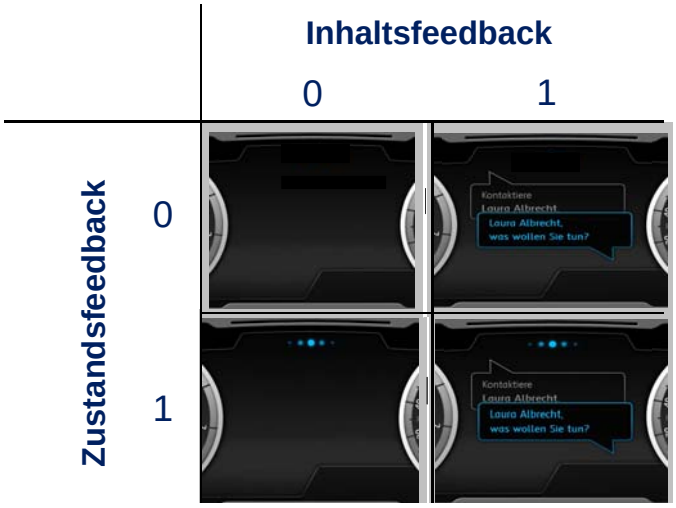

Abbildung 29: Versuchsplan Studie I

Das folgende Kapitel führt die verwendeten Erhebungsinstrumente ein.

6.1.2.4 Erhebungsinstrumente

Zur Erfassung der allgemeinen Gebrauchstauglichkeit fanden im Wesentlichen zwei standardisierte Usability Fragebögen Anwendung; die System Usability Scale und das Software Usability Measurement Inventory. Die **System Usability Scale** (Brooke, 1996; SUS), die entlang der DIN EN ISO-Norm 9241-11 entwickelt wurde, umfasst zehn Einzelitems, die auf einer fünfstufigen Likert-Skala bewertet werden sollten. Für die vorliegenden Untersuchungen wurde die Skalenbeschriftung der SUS leicht adaptiert, sodass anstelle der alleinigen Beschriftung der Pole mit „lehne stark ab“ und „stimme stark zu“, eine kontinuierliche Kodierung durch Plus- und Minus-Zeichen (siehe Abbildung 30) gegeben war.

--	-	0	+	++

Abbildung 30: Skalenbezeichnung SUS

Weiterhin wurde eine deutsche Übersetzung der Items vorgenommen, um sprachbedingte Antwortschwierigkeiten zu vermeiden und das vierte Item durch ein Beispiel erweitert. Im Anhang 12.2.4 kann die Gegenüberstellung der Original-Items mit der Übersetzung nachvollzogen werden.

In Anlehnung an Bangors et al. (2008) wurde der SUS-Fragebogen durch weitere Erhebungsinstrumente ergänzt, da er nur bedingt diagnostischen Schlüsse auf die Ursache der Bewertung

zulässt. Um beispielsweise verschiedene Aspekte der nutzerseitig wahrgenommenen **Systemtransparenz** zu adressieren, wurden sechs Items der SUS-Skala angeschlossen, die der Tabelle 12 entnommen werden können.

Tabelle 12: Systemtransparenzskala

Dimension	**Item**
Handlungsoptionen	*Ich wusste immer, was ich tun kann.*
Ablaufende Prozesse	*Ich wusste immer, was das System tut.*
Verständnisevidenz	*Ich wusste immer, was das System verstanden hat.*
Anzeige Systemausgabe	*Ich wusste immer, was das System gesagt hat.*
Systemgrenzen	*Ich bin mir der Grenzen des Systems bewusst.*
Systemkompetenzen	*Ich habe ein Gefühl dafür, was das System kann.*

Mit diesen selbstkonstruierten Items wurde einerseits erfragt, inwiefern die Nutzer über die verfügbaren Handlungsmöglichkeiten und ablaufenden Prozesse informiert waren. Andererseits wurde die Frage adressiert, ob das System ausreichend Verständnisevidenzen lieferte und die Anzeige der Systemausgabe nützlich war. Das letzte Itempaar sollte das Ausmaß erfassen, in dem Nutzer sich der Systemkompetenzen und -grenzen bewusst waren. Die Skalenbezeichnung wurde an die SUS-Skala angepasst.

Ergänzend wurde die deutsche Version des **Software Usability Measurement Inventory** (SUMI) von Kirakowski und Corbett (1993) erhoben. Der Fragebogen umfasst 50 positiv und negativ formulierte Items, die auf einer dreistufigen Skala beurteilt werden müssen (stimme zu, stimme nicht zu, weiß nicht). Die Items können fünf Dimensionen zugeordnet werden, deren Beschreibung Anhang 12.2.4 entnommen werden kann. Aus den fünf Subskalen lässt sich ein Gesamtwert errechnen, der die allgemeine Gebrauchstauglichkeit beschreiben soll. Der Fragebogen wurde in Ergänzung zum SUS erhoben, da er zusätzlich zu den Dimensionen der Gebrauchstauglichkeit auch die Steuerbarkeit des Systems (Control), die Selbstbeschreibungsfähigkeit (Helpfulness) und die Lernförderlichkeit (Lernability) erfasst und damit auch Aspekte der DIN EN ISO-Norm 9241-110 (Kapitel 2.1.3) abdeckt. Die Affektskala kann dabei als Zufriedenheitskonstrukt herangezogen werden.

Zusätzlich zu diesen subjektiven Beurteilungen der Gebrauchstauglichkeit fand eine objektive Erhebung der Usability-Dimensionen Effektivität und Effizienz statt. Den **objektiven Maßen** lagen Aufzeichnungen des Blickerfassungssystems (Dialogdauer) und Protokollmaße (Fehlbedienung & -erkennung) zugrunde.

Neben der Aufzeichnung der Fahr- und Blickdaten erfolgte die subjektive Einschätzung der Ablenkung in Anlehnung an den **Driving Activity Load Index** (Pauzié, 2007, 2008; Pauzié & Pachiaudi, 1997; DALI), der eine an die Gegebenheiten des Fahrtkontext adaptierte Version des NASA Task Load Index (Hart & Staveland, 1988; NASA-TLX) darstellt. Grundlegende Annahme des DALI ist, dass die Belastung während der Fahrt durch sechs Dimensionen bedingt wird, die in Tabelle 13 dargestellt werden.

Tabelle 13: Dimensionen der Belastung

Dimension	**Zur Erfassung…**
Anforderung an die Aufmerksamkeit	…der Aufmerksamkeit, die von der Fahraufgabe erfordert wird (entscheiden, denken, wählen, etc.).
Visuelle Anforderung	…der visuellen Beanspruchung, die durch die Fahraufgabe entsteht.
Auditive Anforderungen	…der auditiven Beanspruchung, die durch die Fahraufgabe entsteht.
Zeitliche Anforderungen	…des zeitlichen Drucks durch die Fahraufgabe.
Beeinträchtigung der Fahraufgabe	…von möglichen Interferenzen einer Zweitaufgabe mit der Fahraufgabe.
Situationsbezogene Belastung	…verschiedener Belastungen während der Fahrt (Müdigkeit, Unsicherheit, Stress, etc.).

Für die vorliegende Studie wurde die Antwortskala leicht adaptiert, sodass kongruent zu der SUS-Skala eine Bewertung auf einer fünfstufigen Likert-Skala vorgesehen wurde.

Um die Verständlichkeit der Items zu erhöhen und damit die unterschiedliche Beurteilung der Probanden aufgrund ihrer individuellen Interpretation zu minimieren, wurden die Items eindeutiger und umgangssprachlicher formuliert. So wurde jede Dimensionsbeschreibung in einen Ich-Aussagesatz transformiert, der die Versuchspersonen persönlich involvierte. Darüber hinaus wurden die Dimensionen der visuellen und auditiven Anforderungen auf die spezifische Zweitaufgabe angepasst. So sollten die Versuchspersonen nicht, wie vorgesehen, die visuelle oder auditive Beanspruchung der Fahraufgabe beurteilen, sondern den Ablenkungsgrad der Zweitaufgabe. Die auditive Beanspruchung wurde dabei mit der Sprachausgabe gleichgesetzt, die visuelle Beanspruchung mit dem Ablenkungseffekt der Anzeige. Dem Anhang 12.2.4 kann die Erhebung der Belastung in Anlehnung an den DALI entnommen werden.

Weiterhin sollten die Feedbackanzeigen im Rahmen eines **strukturierten Interviews** bewertet werden. Nach einer allgemeinen Einschätzung der Anzeige sollten die Eigenschaften Verständlichkeit, Größe, Lesbarkeit, Attraktivität und Position der Anzeige bewertet werden. Je nach zugewiesener Bedingung sollte auch die Gestaltung des Zustandsfeedbacks (Punktereihe)

und/oder des Inhaltsfeedbacks (Sprechblasen) bewertet werden[21]. Eine erschöpfende Übersicht der Original–Items ist dem Anhang 12.2.4 zu entnehmen.

Die Antwortkategorien wurden in Anlehnung an das von Heller (1985) zur Hörfeldaudiometrie konzipierte **Kategorienunterteilungsverfahren** (im Folgenden: KU-Skala) designt (siehe Abbildung 31). Das Vorgehen ermöglichte die differenzierte Erfassung absoluter Eigenschaftsurteile (Vollrath & Krems, 2011) und die Annahme des Intervallskalenniveaus. Nach Vollrath und Krems (2011) bietet dieses Verfahren auch den Vorteil, dass die Verbalkategorien stark der Umgangssprache der Probanden entsprechen und ihnen somit eine Beurteilung der verschiedenen Stufen leicht fällt. Eine Abbildung dieser Skala lag den Probanden bei der Durchführung des Interviews vor.

sehr schlecht			schlecht			mittel			gut			sehr gut		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Abbildung 31: KU-Skala

Im Rahmen eines anschließenden **Rekognitionstests** wurden die Erinnerungsleistung und die mentalen Modelle[22] der Versuchspersonen erfragt. Nach Beendigung der Aufgaben wurden den Versuchspersonen je nach Versuchsbedingung die entsprechenden Anzeigen des Sprachdialogsystems noch einmal in randomisierter Reihenfolge aufgeschaltet. Dabei sollten sie sich an die Funktion der jeweiligen Anzeige im Dialogablauf erinnern und konnten aus vorgegebenen Antworten wählen. Beendet wurde das Interview durch eine Reihe offener Fragen, bei dem der Versuchsleiter unter anderem nach unklaren, fehlenden oder überflüssigen Zuständen fragte.

Als abschließende Erhebung sollten sich die Probanden an so viele Eingabekommandos wie möglich erinnern und diese benennen. Dies diente der objektiven Erfassung des Wissens der Versuchspersonen über die Systemkompetenzen. Darüber hinaus sollten die Probanden die Nützlichkeit der einzelnen Visualisierungen beurteilen.

Eine Übersicht aller Erhebungsinstrumente liefert der Versuchsablauf in Anhang 12.2.4.

---

[21] Innerhalb der Kontrollgruppe entfielen alle visualisierungsspezifischen Erhebungen.

[22] Mentale Modelle können als vereinfachte Abbilder persönlicher Erfahrungen beschrieben werden und dienen als Orientierung und Bewertungsgrundlage in unbekannten Interaktionssituationen.

6.1.2.5 Stichprobe

Für die Versuchsteilnahme wurde der Besitz eines Führerscheins sowie ungestörtes Farbsehvermögen vorausgesetzt. Entsprechend dieser Kriterien wurden 43 Personen aus der Datenbank des betriebsinternen Probandenpools rekrutiert. Somit nahmen am Versuch ausschließlich Mitarbeiter der Volkswagen AG teil. Alle Probanden waren kompetent im Umgang mit der deutschen Sprache in Wort und Schrift. Die Altersspanne lag zwischen 20 bis 57 Jahren (*MW*= 36.5 Jahre; *SD*= 9.5 Jahre).

Mit 12 weiblichen und 31 männlichen Probanden wurde zwar keine Gleichverteilung der Geschlechter erreicht, allerdings wurde der Anteil weiblicher Probanden gleich auf die Versuchsbedingungen verteilt. Tabelle 14 stellt die Verteilung der Stichprobeneigenschaften auf die Zwischensubjektbedingungen dar.

Tabelle 14: Stichprobenaufteilung nach Feedbackbedingung

		Alter				Geschlecht	
	N	***Minimum***	***Maximum***	***Mittelwert***	***SD***	***männlich***	***weiblich***
Kontrolle	11	22	57	36.8	9.7	8	3
Zustand	11	21	55	38.6	11.3	8	3
Inhalt	11	25	50	37.9	8.1	8	3
Zustand & Inhalt	10	20	48	32.4	8.6	7	3

6.1.2.6 Durchführung

Vor Studienbeginn wurden die Versuchsteilnehmer informiert, dass es sich bei der geplanten Untersuchung um einen Simulatorversuch zur Bewertung eines neuartigen Sprachdialogsystems zur Telefon- und Navigationsbedienung handelte. Im Rahmen der standardisierten Instruktion (Anhang 12.2.2) wurden sie auch über die Aufzeichnung von Fahr- und Blickdaten in Kenntnis gesetzt.

Es folgte die soziodemografische Befragung, in der zusätzlich Erfahrungswerte mit Sprachapplikationen und das Bildungsniveau erfasst wurden. Im Anschluss daran wurden alle Teilnehmer standardisiert für Fahr- und Bedienaufgabe instruiert und das Blickerfassungssystem kalibriert.

Innerhalb der Instruktion in das Sprachdialogsystem erfolgte keine explizite Erklärung der Visualisierungen. Allerdings gab das Beispielkommando *Ich möchte Julia Sommer mobil anrufen.,* das

allen Probanden im Stand vorgeführt wurde, einen ersten Einblick in das eingestellte Rückmeldekonzept. Jeder Teilnehmer wurde darauf hingewiesen, dass ihm im Laufe der Untersuchung Kommandos, aber auch Szenarien vorgegeben werden. Alle Probanden wurden explizit instruiert, dass sie die Fahraufgabe so wenig wie möglich vernachlässigen sollten. Jeder Teilnehmer absolvierte eine Eingewöhnungsfahrt, eine Aufgabenfahrt, sowie eine Referenzfahrt vor der Aufgabenfahrt. Während der Aufgabenfahrt führten die Probanden 12 Interaktionen mit dem SDS aus den Bereichen Telefon und Navigation innerhalb einer der vier Feedbacklösungen aus. Die Erhebung dauerte durchschnittlich 60 Minuten. Ihr schematischer Ablauf kann Anhang 12.2.1 entnommen werden.

6.1.2.7 Rahmenbedingungen der Auswertung

Alle Analysen erfolgten mit einem **Signifikanzniveau** von $\alpha = .05$. Ein Trend wird immer dann berichtet, wenn $p < .10$. Neben dem Signifikanzniveau werden auch **Effektstärken** berichtet. Als Maß der praktischen Bedeutsamkeit wird bei allen Studien, aufgrund der mehrfaktoriellen Designs, das partielle η^2 berichtet.

Im Rahmen der gesamten Arbeit wird sich bei der Interpretation von multivariaten Varianzanalysen auf die **Prüfgröße** Hotelling-Lawleys Spurkriterium bezogen, da der Fokus der Testung von Gruppenunterschieden auf einer gemeinsamen Dimensionen der abhängigen Variablen (wie z. B. Effizienz, Effektivität, etc.) liegt und nicht auf einzelnen abhängigen Variablen. Bei dem Betrachten eines gemeinsamen Konstrukts, zeigt das Hotelling-Lawleys Spurkriterium eine hohe Robustheit bei der Verletzung der Voraussetzungen auf (Wolf, 2003). Da in der ersten Erhebung beide Faktoren nur 2 Stufen aufweisen, führen alle vier Prüfgrößen zu demselben Ergebnis. In den folgenden Untersuchungen wird sich auf den Bericht des Hotelling-Lawleys Spurkriterium beschränkt.

Zur **Prüfung der Voraussetzungen** wurde sich an Bortz (2005) und Wirtz und Nachtigall (2006) angelehnt. Für die Kategorienunterteilungsskala nach Heller (1985) und die kontinuierliche gelabelten Likert-Skalen der Gebrauchstauglichkeits-, Belastungs- und Systemtransparenzbewertungen kann **Intervallskalenniveau** angenommen werden. Auch für die protokollierten Fehlerhäufigkeiten innerhalb der Nutzer-System-Interaktion kann dieses Skalenniveau vorausgesetzt werden.

Selbst bei einer Verletzung der **Normalverteilungsannahme** der betreffenden Variablen kann davon ausgegangen werden, dass der F-Test bei den gegebenen gleichen Zellengrößen robust gegenüber nicht-normalverteilten Rohwerten ist (vgl. Everitt, 1996). Durch die Einhaltung von balancierten Versuchsdesigns ist dies stets gewährleistet. Weiterhin kann die in der vorliegenden Untersuchung adressierte zweifaktorielle Fragestellung, mit non-parametrischen Tests nur inef-

fektiv ausgewertet werden. Es wird sich daher auf den Bericht der multi- bzw. univariaten Varianzanalysen beschränkt.

Ebenso wurde der Levene-Test durchgeführt, um **Varianzhomogenität** im univariaten Fall zu überprüfen. Da heterogene Fehlervarianzen den F-Test bei gleichgroßen Zellenbesetzungen nur unerheblich beeinflussen (Wirtz & Nachtigall, 2006), fand im Rahmen der drei Hauptuntersuchungen keine Anpassung des alpha-Niveaus statt. Im multivariaten Fall erfolgte die Prüfung der **Homoskedastizität** über den Box's-M Test. Auch im diesem Fall sind die Auswirkungen nicht vorhandener Homoskedastizität unerheblich, sofern die Zellenbesetzung annähernd gleich ist. Innerhalb der Analysen der drei Hauptstudien kann der Box's M-Test damit vernachlässigt werden (Tabachnick & Fidell, 2001) und wird nicht berichtet.

Weiterhin wurde die **Multikollinearität** der erklärenden Variablen im Falle einer Regression überprüft. Im Falle einer zu hohen Korrelation zwischen den unabhängigen Variablen (d.h. Kor[Y1,Y2]> .7) wurden die redundanten Variablen entfernt.

Im Folgenden sind die Ergebnisse der Analysen aufgeführt.

6.1.3 Ergebnisse

Das 2x2 Untersuchungsdesign (siehe Abbildung 29) wurde je nach Anzahl der abhängigen Variablen durch eine ANOVA oder MANOVA mit den beiden Zwischensubjektfaktoren (Zustands- und Inhaltsfeedback) als unabhängige Variable ausgewertet. Im Folgenden soll die Prüfung der Hypothesen dokumentiert werden. Zunächst werden die Usability-spezifischen Ergebnisse dargestellt.

6.1.3.1 Gebrauchstauglichkeit und Systemtransparenz

Der Gesamtscore der **System Usability Scale** lag in allen vier Versuchsbedingungen über 70 Punkten. Trotz des hohen Niveaus in allen Bedingungen zeigte sich ein Trend für das Zustandsfeedback (F[1,39]= 3.4; p= .07; η^2= .08) und eine signifikante Interaktion der beiden Faktoren (F[1,39]= 4.4; p= .04; η^2= .10). Es konnte kein Haupteffekt für das IF festgestellt werden (F[1,39]= 1.1; p= .29).

Während eine alleinige Darbietung des Zustands- oder Inhaltsfeedback zu einem Abfall des Gesamtwertes führte, konnte eine kombinierte Darbietung beider Feedbackarten zu einer leichten Verbesserung des SUS-Scores führen verglichen zur jeweils einzelnen Darbietung (siehe Abbildung 32). Bei der vorliegenden disordinalen Interaktion kann der Haupteffekt des Zustandsfeedbacks nicht interpretiert werden.

Abbildung 32: SUS-Score (1 = Minimalwert Usability 100 = Maximalwert) in Abhängigkeit des Inhalts- und Zustandsfeedbacks.

Für den Gesamtscore des **SUMI** zeigte sich ein tendenzieller Haupteffekt des Zustandsfeedback. So fiel der Gesamtscore des SUMI signifikant ab, wenn die Systemzustände vermittelt wurden (F[1,39]= 4.1; p= .05; η^2= .09). Es konnte kein Haupteffekt für das IF festgestellt werden (F[1,39]= 1.1; p= .30). Um die detaillierten Einflüsse der Zwischensubjektfaktoren auf die einzelnen Dimensionen der Beanspruchung zu identifizieren, wurden die fünf Subskalen des SUMIs mit einzelnen Varianzanalysen analysiert. Dabei zeigten sich Effekte des Zustandsfeedbacks auf die Dimensionen Kontrollierbarkeit, Erlernbarkeit und Affekt (siehe Tabelle 15). Durch Hinzunahme des Zustandsfeedbacks kam es stets zu einer Verschlechterung der Beurteilung.

Tabelle 15: Effekte des Zustandsfeedbacks auf die SUMI-Skalen

	Effekt Zustand		
	$F_{(1,39)}$	p	η^2
Affekt	2.9	.10	.07
Kontrollierbarkeit	**4.5**	**.04**	**.10**
Erlernbarkeit	**4.2**	**.05**	**.10**
Effizienz	1.8	.19	
Selbstbeschreibungsfähigkeit	2.4	.13	

Es konnten keine Haupteffekte des Inhaltsfeedbacks (F[1,39]= 1.1; p= .30) oder Interaktionseffekte der Feedbackarten (F[1,39]= 1.0; p= .34) festgestellt werden.

Für die globale Analyse der **Systemtransparenz** ergaben sich weder Haupt- noch Interaktionseffekte (siehe Tabelle 16).

Tabelle 16: Feedbackeffekte Systemtransparenz

	Effekt Zustand		Effekt Inhalt		Effekt Zustand*Inhalt	
	$F_{(6,34)}$	p	$F_{(6,34)}$	p	$F_{(6,34)}$	p
Systemtransparenz	0.7	.63	0.9	.49	1.1	.39

Bei der heuristischen Analyse der Einzelitems zeigte sich eine Wechselwirkung bei dem Item *Ich wusste immer was das System verstanden hat.* (F[1,39]= 4.4; p= .04; η^2= .10) (siehe Abbildung 33).

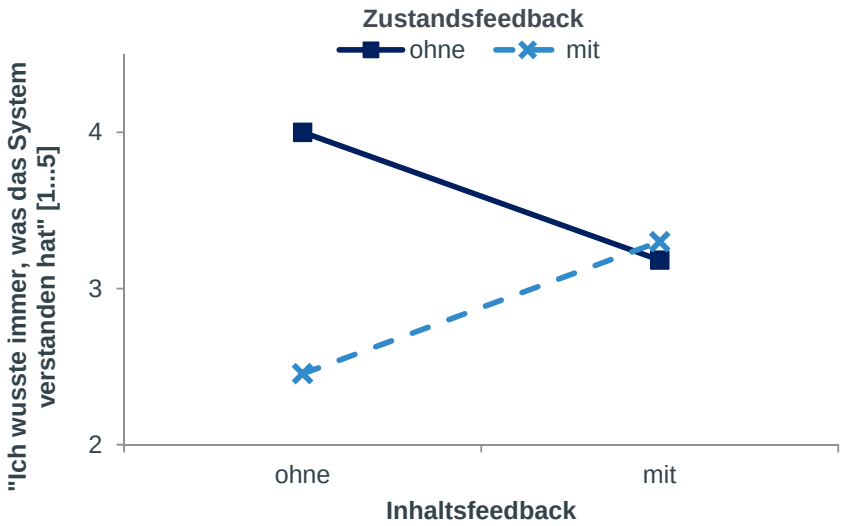

Abbildung 33: Systemtransparenz (1 = sehr wenig 5 = sehr viel) in Abhängigkeit des Inhalts- und Zustandsfeedbacks.

Während eine alleinige Darbietung des Zustandsfeedbacks zu einer Verringerung der Zustimmung führt, kann eine kombinierte Darbietung beider Feedbackarten zu einer leichten Verbesserung der Beurteilung führen.

6.1.3.2 Effektivität, Effizienz und Zufriedenheit

Zur Betrachtung des Einflusses der Feedbackarten auf Fehlerkennungen und -bedienungen wurden univariate Analysen der Effektivitätsindikatoren angeschlossen (siehe Tabelle 17).

Tabelle 17: Effekte des Zustandsfeedbacks auf die Effektivitätsindikatoren

	Effekt Zustand		**Effekt Inhalt**		
	$F_{(1,39)}$	p	$F_{(1,39)}$	p	η^2
TTE	0.02	.90	**4.9**	**.03**	**.11**
Fehlerkennung	1.90	.18	0.03	.85	
Subjektive Zuverlässigkeitsbeurteilung	2.47	.12	0.84	.37	

Es zeigte sich ein Effekt des Inhaltsfeedbacks auf die Anzahl der zu frühen Spracheingaben. So ließen sich signifikant mehr TTE-Fehler beobachten, wenn das Inhaltsfeedback präsentiert wurde. Es konnten allerdings keine Effekte des Zustandsfeedbacks oder Interaktionseffekte (F[1,39]< 1; p= .95) auf die Anzahl der TTE-Fehler dokumentiert werden.

Insgesamt konnte festgestellt werden, dass innerhalb aller Versuchsbedingungen ein hohes Niveau an Fehlbedienungen und Systemfehlern vorlag. Gewichtet an den Kommandos zeigte sich, dass es durchschnittlich bei jedem Kommando zu 0.5 Substitutionsfehlern und zu 0.2 Fehlerkennungen kam. Die Anzahl der TTE-Fehler näherte sich gemittelt über die Feedbackbedingungen 10 an, was einer Fehlerquote von über 80% pro Kommando entsprach. Diese hohe Fehleranfälligkeit spiegelte sich auch bei der Identifikation problematischer Kommandos wider. Als problematisch wurde ein Kommando immer dann bezeichnet, wenn ein Systemfehler (Fehlerkennung oder Substitutionsfehler) oder eine Fehlbedienung (TTE) auftrat. Insgesamt zeigte sich über den Verlauf ein sehr hohes Level an problematischen Kommandos. Während das erste Kommando bei 79.1% der Probanden einen Fehler beinhaltete, waren es bei dem zweiten Kommando 44.2% und bei dem dritten Kommando 60.5%. Die Anzahl der Fehler war dabei stark abhängig von der Mindestanzahl der Turn Takes. So konnte bei dem siebten Kommando, welches aufgrund des mehrdeutigen Namens Linda mindestens drei Sprecherwechsel benötigte keine fehlerfreie Interaktion aufgezeichnet werden.

Auch bei der detaillierten Betrachtung der manipulierten Kommandos (Kommando 4, 7 und Szenario 12), in denen durch eine Sprechpause ein TTE-Fehler bewusst provoziert wurde, zeigte sich kein Effekt der Feedbackarten (ZF F[1,39]= 1.10; p=.30; IF F[1,39]< 1; p=.66; ZF*IF F[1,39]< 1; p=.98).

Bei der univariaten Analyse der **Effizienzvariablen** konnte lediglich ein Interaktionseffekt der beiden Zwischensubjektfaktoren für die Durchführungsdauer festgestellt werden (ZF*IF: F[1,39]= 3.9; p=.06; η^2=.09). Während die alleinige Vermittlung der Dialoginhalte zu einer Zunahme der Dauer führte, ließ sich eine leichte Abnahme der Dauer im Vergleich zur jenem Einzelfeedback beobachten, wenn zusätzlich zu dem Inhalts- auch Zustandsfeedback vermittelt wurde (siehe Abbildung 34).

Abbildung 34: Bediendauer (in Sekunden) in Abhängigkeit des Inhalts- und Zustandsfeedbacks.

Zur detaillierten Analyse der **Zufriedenheit** fand eine univariate Varianzanalyse mit der SUMI Affektskala statt. Wie bereits in Kapitel 6.1.3.1 berichtet, zeigte sich ein Trend bezüglich des Zustandsfeedbacks (F[1,39]= 2.9; p= .10; η^2= .07).

6.1.3.3 Mentale Beanspruchung und Systemkompentenzeinschätzung

Um die detaillierten Einflüsse der Zwischensubjektfaktoren auf die einzelnen Dimensionen der **Beanspruchung** zu identifizieren, wurden die Faktoren des DALIs in eine MANOVA einbezogen.

Es konnte ein Haupteffekt des Zustandsfeedbacks auf das Gesamtkonstrukt der Beanspruchung ausgemacht werden (F[6,32]= 2.4; p= .05; η^2= .31). Dieser globale Effekt ließ sich auf univariate Effekte des Zustandsfeedbacks auf die Dimensionen Aufmerksamkeitsanforderungen, auditive Anforderungen, situationsbezogene Belastung und Beeinträchtigung der Fahraufgabe zurückführen, die in Tabelle 18 dargestellt sind.

Tabelle 18: Einfluss des Zustandsfeedbacks auf die mentale Beanspruchung

	Effekt Zustand		
	$F_{(1, 37)}$	p	η^2
Ich fühlte mich in der Bewältigung meiner Fahraufgabe beeinträchtigt.	3.9	.06	.10
Ich fühlte mich abgelenkt.	**6.4**	**.02**	**.15**
Die Sprachausgabe hat mich abgelenkt.	**5.2**	**.03**	**.12**
Ich habe mich gestresst gefühlt.	**7.8**	**.01**	**.17**

Dabei zeigte sich durch Hinzunahme des Zustandsfeedbacks stets ein Beanspruchungsanstieg. Abbildung 35 verdeutlicht exemplarisch den subjektiv empfundenen Beeinträchtigungsanstieg durch Hinzunahme des Zustandsfeedbacks.

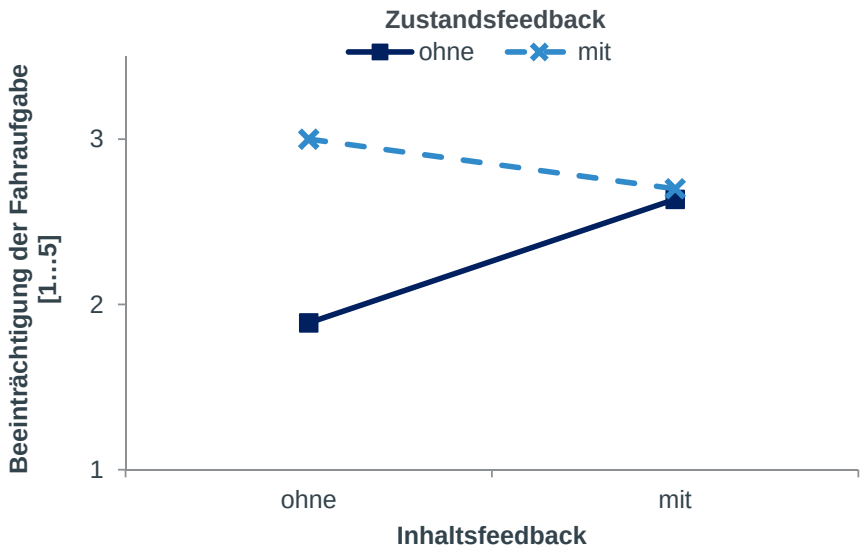

Abbildung 35: Subjektive Beeinträchtigung (1 = sehr wenig 5 = sehr viel) der Fahraufgabe in Abhängigkeit des Inhalts- und Zustandsfeedbacks.

Ferner konnte ein Effekt des Inhaltsfeedbacks auf die Dimension visuelle Anforderung festgestellt werden. Auch hier wurde dem Item *Die visuelle Anzeige hat mich abgelenkt.* häufiger zugestimmt, wenn die Dialoginhalte dargeboten wurden (F[1,37]= 4.2; p= .04; η^2= .11).

Es ergab sich ein Trend bei der Interaktion[23] der Faktoren für das Item *Ich fühlte mich in der Bewältigung meiner Fahraufgabe beeinträchtigt.*(F[1,32]= 3.1; p= .09; η^2= .08). Während die alleinige Anzeige von Inhaltsfeedback zu einem Anstieg der Beeinträchtigung der Fahraufgabe führte, konnte sie bei gegebenem Zustandsfeedback eine Reduktion der Beeinträchtigung verglichen zu dem Einzelfeedback erlangen.

Zur Analyse des Einflusses der Zwischensubjektfaktoren auf das Wissen der Versuchspersonen um die **Systemkompetenzen** wurde die Gesamtanzahl erinnerter Kommandos in eine univariate Varianzanalyse einbezogen. Es zeigten sich weder signifikante Haupt- noch Interaktionseffekte (siehe Tabelle 19).

Tabelle 19: Effekte auf die Anzahl der erinnerten Kommandos

	Effekt Zustand		Effekt Inhalt		Effekt Zustand*Inhalt	
	$F_{(1,38)}$	p	$F_{(1,38)}$	p	$F_{(1,38)}$	p
Kommandorecall	1.1	.31	0.1	.75	1.1	.31

In allen vier Bedingungen lag die Anzahl erinnerter Kommandos mit durchschnittlich 3.1 bis 3.6 von fünf inhaltlich unterscheidbaren Kommandos recht hoch.

6.1.3.4 Visualisierungsinterview

Anschließend wurde das Konstrukt **Gestaltung** analysiert, welches sich aus einer Bewertung der Verständlichkeit, der Größe, der Lesbarkeit, der Attraktivität und einer allgemeinen Anzeigenbewertung zusammensetzte. Die Kontrollgruppe wurde von diesen Analysen ausgeschlossen, da ihnen keine Visualisierung präsentiert wurde. Um Unterschiede zwischen den Umsetzungen zu testen, wurde nur die Gruppe als Faktor (IF, ZF, ZF+IF) herangezogen.

Bei den univariaten Analysen der Gestaltungsvariablen konnte ein signifikanter Gruppeneffekt bei der Bewertung der Größe, der Lesbarkeit und der Attraktivität (siehe Tabelle 20). Dabei waren Attraktivität und Größe stark interkorreliert (r=.7). Weiterhin konnte ein Trend der allgemeinen Bewertung (siehe Tabelle 20) identifiziert werden. Nachfolgende Kontrastanalysen ergaben, dass sich diese vier Variablen signifikant zwischen der Zustands- und Inhaltsfeedbackgruppe unter-

[23] Da es sich um eine hybride Interaktion handelt ist der Haupteffekt des Zustandsfeedbacks weiterhin interpretierbar.

schieden. So schätzen die Versuchspersonen das Inhaltsfeedback auf diesen Gestaltungsratings besser ein. Die Analysen ergaben darüber hinaus, dass auf den Größen- und Attraktivitätseinschätzungen die kombinierte Darstellung von Inhalts- und Zustandsfeedback besser eingeschätzt wurde, als die alleinige Darbietung des Zustandsfeedbacks.

Tabelle 20: Gruppeneinfluss auf die Gestaltungsdimensionen

	Effekt Feedbackgruppe		
	$F_{(2,28)}$	p	η^2
Allg. Bewertung	2.6	.09	.16
Größe	**3.9**	**.03**	**.22**
Lesbarkeit	**3.7**	**.03**	**.21**
Attraktivität	**5.4**	**.01**	**.28**

Es zeigte sich ebenfalls ein deskriptiver Unterschied in der **Gestaltungsbewertung** der Anzeigen (siehe Tabelle 21). So bewertete die Gruppe, die nur Zustandsfeedback sah, die Gestaltung der Anzeige als *mittel*, während die Inhaltsfeedbackgruppe ihre Gestaltung als *gut* bezeichnete. Die kombinierte Gruppe bewertete beide Anzeigen gleichermaßen als *gut*.

Tabelle 21: Gestaltungsbewertung

Gruppe	***N***		***MW***	***SD***
Zustandsfeedback	11	Gestaltung Punktereihe	9.2	1.4
Inhaltfeedback	11	Gestaltung Sprechblasen	11.6	2.6
Zustands- +Inhaltsfeedback	10	Gestaltung Punktereihe	11.5	2.6
		Gestaltung Sprechblasen	10.3	3.1

Die Bewertung des Zustandsfeedbacks unterschied sich zwischen den Gruppen signifikant. So bewertete die Gruppe, die ausschließlich Zustandsfeedback sah, dieses sign. schlechter, als die Gruppe, die beide Feedbackarten präsentiert bekamen (T= -2.6; df= 19; p= .02). Die Bewertung des Inhaltsfeedbacks unterschied sich nicht zwischen den Gruppen (T= 1.1; df= 19; p= .30).

Der angeschlossene **Rekognitionstest** soll nur deskriptiv Erwähnung finden, da eine Vergleichbarkeit zwischen den Gruppen aufgrund unterschiedlicher Feedbackmengen kaum gewährleistet war. Während die kombinierte Gruppe sich an fünf Rückmeldungsarten erinnern musste, waren es bei der Inhaltsfeedbackgruppe lediglich drei. Zur Darstellung wird sich auf eine kategorienba-

sierte Auswertung der Antworten der Versuchspersonen (richtig oder falsch) bezogen. In Tabelle 22 lässt sich der prozentuale Anteil richtiger Antworten ablesen.

Es zeigte sich, dass die Rekognitionsraten für das Inhaltsfeedback sehr hoch ausfallen. Das Zustandsfeedback wurde nur dann in einem akzeptablen Ausmaß erinnert, wenn es kombiniert zu den Dialoginhalten angezeigt wurde. Verwechslungen traten insbesondere zwischen dem Zuständen Ready und Processing auf.

Tabelle 22: Prozentualer Anteil richtiger Antworten

Gruppe	N	Ready	Processing	User result	System speaking understood	System speaking not und.
Zustandsfeedback	11	27.3 (3/11)	54.5 (6/11)	/	45.5 (5/11)	54.5 (6/11)
Inhaltsfeedback	11	/	/	81.8 (9/11)	81.8 (9/11)	90.9 (10/11)
Zustandsfeedback-+Inhaltsfeedback	10	80 (8/10)	80 (8/10)	90 (9/10)	70 (7/10)	90 (9/10)

Auch die **Nützlichkeitsbewertung** des Zustands- und Inhaltsfeedbacks unterschieden sich zwischen den Gruppen deskriptiv. So wurde die Nützlichkeit der Zustandsanzeigen im mittleren Bewertungssegment verortet. Die Nützlichkeit des Inhaltsfeedback wurde weitestgehend als *gut* bezeichnet. Allein die Darstellung der eigenen Spracheingabe bezeichneten die VP als weniger hilfreich. Bei den kombinierten Darstellungen wurde die Nützlichkeit aller Rückmeldungen als *gut* eingestuft.

Die Kontrollgruppe wurde ausschließlich mit der Frage konfrontiert, ob sie sich eine unterstützende Visualisierung wünschen würden. 70% der Probanden, die weder Zustands- noch Inhaltsfeedback dargeboten bekamen, bejahten diese Frage und wünschten sich eine grafische Repräsentation der Empfängerzustände des Sprachdialogsystems.

6.1.3.5 Blickdaten

Zur Auswertung der Blickabwendung fand eine ANOVA mit Messwiederholung innerhalb der Zwischensubjektfaktoren Inhaltsfeedback und Zustandsfeedback statt. Dabei wurden drei Messzeitpunkte (Baseline, Kommandos, Szenarios) definiert, um einen Anstieg der Blickabwendung während der Aufgabenfahrten zu testen und ggf. Gewöhnungseffekt zu identifizieren. Da die Aufzeichnung der Blickdaten aufgrund technischer Probleme nicht bei jeder Versuchsperson erfolgreich war, belief sich die Stichprobe hier auf 37 Probanden.

Die Blickabwendung wurde definiert durch die prozentuale Anzahl der Blicke auf das Kombidisplay während der Fahrtzeit. Der Anzeigenbereich in der Mitte des Kombiinstruments wurde als Untersuchungsbereich (engl. *area of interest*) definiert. Dadurch lagen auch Teile des Tachografen und der Drehzahlmesser in diesem Bereich und es erfolgte die Kontrolle anhand der Baseline. Blinzler und Durchblicke wurden mit Hilfe der Blickanalysesoftware herausgerechnet.

Es zeigte sich kein signifikanter Effekt der Messzeitpunkte auf die Blickabwendung (F[2,62]= 1.6; p= .21). Dennoch konnte ein Interaktionseffekt des Inhaltsfeedbacks beobachtet werden (F[2,62]= 10.7; p< .0001; η^2= .26), wonach die Anzeige der Dialoginhalte je nach Messzeitpunkt unterschiedlichen Auswirkungen hatte (siehe Abbildung 36). Während im Rahmen der Baselinefahrt (keine Anzeige) nur eine geringe Anzahl Kombiblicke registriert wurde, kam es während der Kommandofahrt zu einem deutlichen Anstieg der Blickabwendung, der sich im Rahmen der Szenariofahrt ein wenig relativierte.

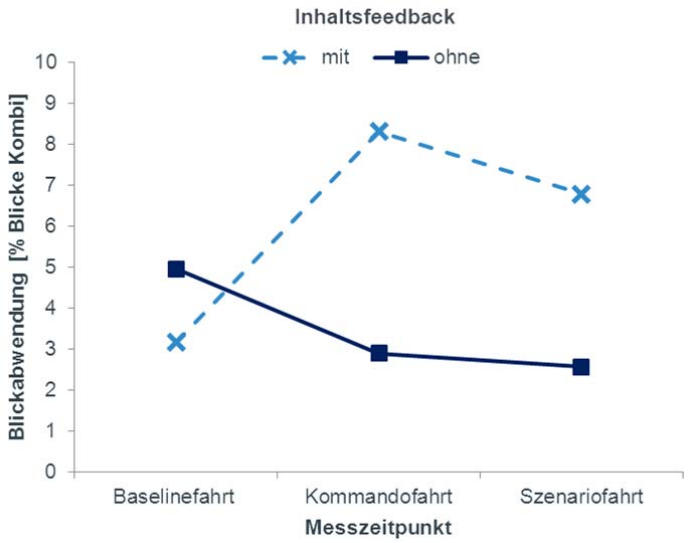

Abbildung 36: Prozentuale Häufigkeit der Blicke auf das Kombidisplay in Abhängigkeit des Inhaltsfeedbacks.

Es zeigte sich somit auch ein Haupteffekt des Zwischensubjektfaktors Inhaltsfeedback (F[1,31]= 5.3; p= .03; η^2= .15), wonach eine Darstellung der Dialoginhalte zu einem Anstieg der Blickabwendung führte. Interaktions- oder Haupteffekte des Faktors Zustandsfeedback ließen sich nicht beobachten (ZF F[1,31]< 1; p= .58; Blickabwendung*ZF F[2,62]= 1.4; p= .25).

6.1.3.6 Fahrdaten

Die Gruppen unterschieden sich hinsichtlich ihrer Fähigkeiten zur Spurhaltung während der Referenzfahrt nicht (ZF $F[1,38]< 1$; $p= .86$; IF $F[1,38]< 1$; $p= .42$). Zur Analyse der Spurhaltung wurde die SDLP als Innersubjektfaktor in einer ANOVA mit Messwiederholung mit den Zwischensubjektfaktoren Inhalt und Feedback betrachtet. Dabei wurden erneut drei Messzeitpunkte definiert (Baseline, Kommandofahrt, Szenariofahrt). Da die Aufzeichnung der Fahrdaten aufgrund technischer Probleme nicht bei jeder Versuchsperson erfolgreich war belief sich die Stichprobe auf 36 Probanden.

Dabei zeigten sich keine Effekte des Innersubjektfaktors. So konnten keine Differenzen der Spurabweichung (SD) über die Messzeitpunkte festgestellt werden ($F[2,31]< 1$; $p= .66$). Ebenso konnten keine Interaktionseffekte der Messzeitpunkte mit den Zwischensubjektfaktoren (ZF $F[2,31]< 1$; $p= .49$; IF $F[2,31]<1$; $p= .43$) und keine Effekte der Zwischensubjektfaktoren Inhalt- ($F[1,32]< 1$; $p= .97$) oder Zustandsfeedback ($F[1,32]< 1$; $p= .91$) dokumentiert werden.

Nachdem die Ergebnisse der Studie nun dargestellt wurden, sollen sie im Folgenden vor dem Hintergrund der ursprünglichen Frage- und Problemstellungen diskutiert werden.

6.1.4 Diskussion

Durch die Unterstützung bei der Etablierung des Common Grounds und der frühen Möglichkeit zur Korrektur, zielte die Umsetzung der visuellen Rückmeldungen primär auf die Effektivität der Dialoggestaltung ab. Sowohl Dialogzustand, als auch die Dialoginhalte sollten stets klar erkennbar sein und zu einer Steigerung der Gesamtusability führen.

Entgegen dieser Erwartung konnte nicht belegt werden, dass die zusätzlichen Visualisierungen die **Gebrauchstauglichkeit** des Sprachdialogsystems erhöhen. Weder Zustands- noch Inhaltsfeedback führten zu einer besseren Bewertung der allgemeinen Usability des Systems. Es zeigte sich sogar bei der alleinigen Vermittlung von Zustandsfeedback ein negativer Effekt auf die Bewertung der allgemeinen Gebrauchstauglichkeit des Systems, verglichen zur Kontrollgruppe. Dieser konnte durch die Präsentation von Inhaltsfeedback tendenziell abgemildert werden. Ohne jegliche Visualisierung (KG) ließ sich jedoch der höchste SUS-Score beobachten.

In ähnlicher Weise zeigte sich dies auch für die Beurteilung der **Systemtransparenz**. So äußerten Probanden, die ausschließlich die Zustände des Systems vermittelt bekamen, nicht immer zu wissen, was das System verstanden hat. Durch die zusätzliche Darbietung von Inhaltsfeedback, konnte auch dieser negative Effekt verkleinert werden.

Beide disordinalen Interaktionen gingen dabei maßgeblich auf den verschlechternden Effekt des Zustandsfeedbacks verglichen zur KG zurück, welcher durch das Inhaltsfeedback ein wenig Abmilderung fand.

Die Ergebnislage wurde durch die Auswertung des SUMI, der sich neben den Kernaspekten der Usability auch der Thematik des **Joy of Use** nähert, gestützt. Insgesamt verschlechterte sich der Gesamtscore des SUMI, die Einschätzung der Kontrollierbarkeit, der Erlernbarkeit und die affektive Systembewertung, wenn die Systemzustände vermittelt wurden. Jene Effekte konnten nicht durch die Hinzunahme von Inhaltsfeedback abgemildert werden.

Damit konnte kein positiver Effekt der Zustandsvermittlung auf die nutzerseitige **Zufriedenheit** nachgewiesen werden. Sogar die bereits berichtete affektive Systembeurteilung zeigte negative Effekte, wenn die Dialogzustände vermittelt wurden.

Hinsichtlich der **Effektivität** des Systems wurde keine Reduktion der Systemfehler oder Fehlbedienungen durch die Darbietung des Zustandsfeedbacks festgestellt. Der unterstützende Effekt durch die Vermittlung zulässiger Interaktionen konnte somit nicht belegt werden. Auch bei der Betrachtung der manipulierten Kommandos, in denen durch eine Sprechpause ein TTE-Fehler provoziert wurde, kann diese Annahme nicht bestätigt werden. Dagegen konnte ein unerwarteter Anstieg der Anzahl zu früher Spracheingaben bei der Darbietung von Inhaltsfeedback beobachtet werden. Dies lässt sich maßgeblich auf ein Problem bei der gestalterischen Umsetzung zurückführen. So konnte der Dialogmanager den vollständigen Systemprompt erst dann an die grafische Schnittstelle weitergeben werden, wenn das System die Sprachausgabe beendet hatte. Ein Auftauchen der (gefüllten) blauen Sprechblase erfolgte also in den letzten Zügen der Systemausgabe. In dieser Zeit war das System zumeist noch nicht wieder empfangsbereit. Die Versuchspersonen missinterpretierten womöglich das Erscheinen der Systemsprechblase als einen Bereitschaftsindikator und formulierten eine neue Eingabe zu früh.

Eine alternative Erklärung für die vermehrte Rate der zu frühen Spracheingaben bei dargebotenem Inhaltsfeedback kann auch sein, dass Nutzer durch die textuelle Wiedergabe bereits zu einem früheren Zeitpunkt darüber informiert waren, ob ihre Spracheingabe richtig verstanden wurde und so durch eine schnelle Eingabe den Dialog effizienter gestalten wollten.

Insgesamt wurde damit bei der Implementierung das relevante Kriterium der zeitgleichen Wiedergabe verletzt, welches nötig ist, um die Vorteile der kombinierten Darbietung von Sprachausgabe und Visualisierung zu erreichen. Im Rahmen des Klassifikationsschemas multimodaler Systeme nach Nigay und Coutaz (1993) konnte nicht wie angestrebt, die parallel-redundante Interaktion realisiert werden. Auf die aufeinanderfolgende Präsentation der gleichen Informationen über verschiedene Modalitäten reagierten die Versuchspersonen empfindlich. Durch diese

(nicht beabsichtigte) sequentiell-redundante Interaktion kann davon ausgegangen werden, dass die positiven Effekte des Inhaltsfeedbacks unterschätzt werden. Bei zukünftigen Umsetzungen muss der Zeitpunkt der Systemausgaben verbessert werden, denn das Zeitintervall zwischen Nutzeraktion und Feedback ist ein ebenso wichtiger Faktor wie das Feedback selbst (vgl. Saffer, 2007, S. 47).

Auch wenn dieser Effekt der Visualisierungen nicht erwartungskonform ausfiel, so konnte damit doch nachgewiesen werden, dass einerseits Nutzer sensibel auf die ergänzenden Anzeigen reagieren und andererseits diese als Signale für Sprecherwechsel fungieren können.

Bei Betrachtung der durchschnittlichen **Dialogdauer** ließ sich ein Effekt beobachten, der sich auf die größere Anzahl der zu frühen Spracheingaben in den Inhaltsfeedbackgruppen zurückführen ließ. So führte die alleinige Vermittlung der Dialoginhalte ohne Zustandsfeedback zu einer Zunahme der Dauer. Dagegen ließ sich eine leichte Abnahme der Dauer feststellen, wenn zusätzlich zu dem Inhalts- auch Zustandsfeedback vermittelt wurde. Die Vermittlung der Systemzustände hatte an dieser Stelle anscheinend einen positiven Effekt auf die Systemeffizienz, da diese zu einer Vermeidung von Fehlbedienungen beitragen konnten. Die negativen Effekte des Sprechblasen-Timings konnten somit durch die Vermittlung der Systemzustände abgemildert werden.

Insgesamt muss die **Effektivität** des SDS als gering eingestuft werden. So wiesen 79.4% der Kommandos eine Fehlerart (TTE, Fehlerkennung oder Substitutionsfehler) auf. Auch wenn einzelne Kommandos drei Sprecherwechsel erforderten und per Manipulation TTE-Fehler provozierten, so bedeutet dies eine inakzeptable Erkennerleistung und eine zu hohe Fehleranfälligkeit des Systems. Dies schlug sich auch in der subjektiven Bewertung des Items *Das System versteht mich gelegentlich unerwartet falsch.* nieder, dem über 60% der Probanden zustimmten. Somit wurden die Grounding-Prozesse mit dem SDS bereits durch eine geringe Effektivität derart erschwert, dass die Vermittlung der Systemzustände und Dialoginhalte keine Verbesserung erzielen konnten. Es kann davon ausgegangen werden, dass das hohe Level an Fehlerkennungen bzw. -bedienungen die Effekte des Feedbacks in den Hintergrund treten ließ. Insbesondere die Darstellung der Dialoginhalte ist bei einer mangelhaften Systemeffektivität irrelevant, da jene direkt an die Erkennergüte geknüpft sind. Solange eine bestimmte Basiseffektivität des Systems nicht gegeben ist, scheint eine Verbesserung der Zufriedenheit kaum möglich (vgl. Kapitel 7.1: Zwei-Faktoren-Theorie der Usability). Zukünftige Untersuchungen sollten die Effekte von Inhalts- bzw. Zustandsfeedback erneut unter Verwendung eines robusteren SDS evaluieren.

Bezüglich der mentalen **Beanspruchung** zeigte sich ausschließlich ein negativer Effekt des Zustandsfeedbacks. So stiegen bei der Präsentation der Systemzustände sowohl die Aufmerksamkeitsanforderungen, als auch die auditive Anforderungen und die situationsbezogene Belastung. Darüber hinaus empfanden die Versuchspersonen eine größere Beeinträchtigung der Fahraufga-

be. Begründen kann man diese Belastungseffekte mit der zu kleinen und schlecht lesbaren Repräsentanz der Zustände. Die zusätzliche Anzeige von Inhaltsfeedback konnte dabei erneut eine leichte Reduktion der Beeinträchtigung erlangen. Abseits der visuellen Beanspruchung, die bei der Darbietung des Inhaltsfeedbacks einen signifikanten Anstieg zeigte, konnte kein Anstieg der allgemeinen Beanspruchung gegenüber der Kontrollgruppe ohne jegliche Visualisierung beobachtet werden. Allerdings zeigte sich auch nicht die erwartete Entlastung durch die Hinzunahme des Inhaltsfeedbacks.

Alle Probanden erinnerten, unabhängig von der Feedbackgruppe, ähnlich viele Kommandos. Es kann somit nicht davon ausgegangen werden, dass die textuelle Vermittlung der Systeminhalte zu einer transparenteren Vermittlung der **Systemkompetenz** führen.

Hinsichtlich der **Gestaltungsbeurteilung** konnte zwar die Annahme bestätigt werden, dass die Verständlichkeit der beiden Anzeigen gleichermaßen als *gut* beurteilt wurde. Dennoch zeigt sich - entgegen der Erwartungen - ein Gestaltungseffekt. So fallen die Einschätzung der Größe, der Lesbarkeit, der Attraktivität und die allgemeine Bewertung der Anzeige für das Zustandsfeedback schlechter aus. Es zeigte sich auch an dieser Stelle, dass die Größen- und Attraktivitätseinschätzungen bei der kombinierten Darstellung von Inhalts- und Zustandsfeedback besser ausfielen, als die alleinige Darbietung des Zustandsfeedbacks. Diese Befundlage wurde auch von der Gestaltungsbewertung der Punktereihe gestützt. Während die Gestaltung der Sprechblasen als *gut* bezeichnet wurde, wurde die Umsetzung der Punktereihe bei alleiniger Darbietung nur als *mittel* eingeschätzt. Der negative Effekt des ZF relativierte sich erneut bei kombinierter Darbietung beider Feedbackarten.

Eine Gegenüberstellung dieser Gebrauchstauglichkeits- und Gestaltungsbewertungen verdeutlicht eine gewisse Konvergenz. In der Zustandsfeedbackgruppe wurde die geringste Einschätzung der Usability und der Gestaltung beobachtet, während eine kombinierte Darbietung beider Feedbackarten diesen Effekt sowohl bei der Gestaltung, als auch bei der Usabilitybeurteilung abmildern konnte. Die gestalterischen Mängel der Umsetzung sollen daher als Ursache für die schlechtere Usability-Einschätzung in Betracht gezogen werden.

Es zeigten sich ebenso maßgeblich schlechtere **Rekognitionsleistungen** der Zustände. Weniger als die Hälfte der Probanden waren dazu in der Lage, die richtige Funktion den Zuständen zuzuweisen. Diese Befundlage kann drei Erklärungen finden:

Einerseits kann die Verständlichkeit der Darbietung in Frage gestellt werden. Diese zeigte jedoch bei der Analyse des Gestaltungskonstrukts keine Unterschiede zu den anderen Feedbackgruppen. Andererseits kann die bereits berichtete schlechte Lesbarkeit die Ursache für diese mangelhafte Memorierungsleistung sein. Bei einer ungenügenden Sichtbarkeit der Anzeige können sich

die Probanden nur schlecht an die einzelnen Zustände erinnern. Diese Argumentation stützt der Befund, dass bei einer kombinierten und damit größeren Darbietung der Feedbackarten auch die Zustände viel besser erinnert wurden (durchschnittlich 80%).

Ebenfalls zeigte sich, dass insbesondere die Zustände Ready und Processing verwechselt wurden. Dies kann neben der ähnlichen Gestaltung der beiden Zustände auch auf einen hoch frequenten Systemfehler zurückgeführt werden. So zeigte das System ein Fehlverhalten und formulierte nach einem Time-out keine erneute Nachfrage. Dies war insbesondere dann kritisch, wenn Nutzer eine zu frühe Spracheingabe produzierte und damit vor der Erkenneröffnung ihren Befehl einsprachen. Das System reagierte dann mit einer Ready-Phase. Wiederholte der Nutzer seine Eingabe nicht, so ging es ohne erneute Rückmeldung in den Ruhe-Zustand. Dies führte dazu, dass viele Nutzer den Ready- als Processing-Zustand missinterpretierten. Denn während das System noch nichts gehört hatte und auf eine Eingabe wartete (Ready), waren die Nutzer sich sicher, dass das System ihre verfrühte Eingabe gehört hat und diese nun verarbeitet (Processing).

Die Rekognitionsraten der Feedbackanzeigen des Inhaltsfeedbacks fielen hingegen sehr hoch aus. Es kann davon ausgegangen werden, dass der Text in den Sprechblasen den Nutzern eine gewisse Hilfestellung bot, die bei dem Zustandsfeedback nicht gegeben war.

Im Rahmen der **Blickdatenanalysen** konnte belegt werden, dass sich die Blickabwendung von der Straße nur bei der textuellen Repräsentation der Dialoginhalte erhöht. Dennoch zeigte sich innerhalb der Inhaltsfeedbackgruppen im Verlauf der Aufgabenfahrten eine Verringerung der Blickabwendung. Auch wenn dieser Rückgang der prozentualen Kombiblicke nicht signifikant ausfiel, deutet dies auf eine Lernkurve hin, die in Untersuchungen mit längerer Interaktionsdauer vertiefend untersucht werden sollte. Es kann angenommen werden, dass Nutzer nach einer längeren Interaktionsphase auch bei textueller Darbietung keine längere Blickabwendung mehr zeigen.

Für das Zustandsfeedback konnte keine erhöhte Blickabwendung beobachtet werden. Dieser Befund könnte zwei Ursachen haben. Einerseits kann angenommen werden, dass die Punktereihe verglichen zu der textuellen Darbietung auch peripher gut wahrnehmbar ist. Aufgrund des Farbwechsels ist eine Blickabwendung kaum nötig, um die verschiedenen Zustände zu identifizieren. Andererseits sollte aufgrund der geringen Memorierungsleistung der Versuchspersonen und der schlechten Lesbarkeit auch die Vermutung diskutiert werden, ob die Versuchspersonen das Zustandsfeedback eventuell gar nicht beachteten.

Bezüglich der Standardabweichung der lateralen Position konnte die Annahme, dass weder Zustands- noch Inhaltsfeedback zu einer Beeinträchtigung der **Fahraufgabe** führten, bestätigt wer-

den. Insgesamt zeigten die Aufgabenfahrten verglichen zu den Referenzfahrten keine Verschlechterung der Fahrleistung und können damit den Vorteil der Sprachbedienung im Fahrtkontext belegen. Somit konnte die These von Salmen (2002) verifiziert werden, dass ergänzende Visualisierungen als unbedenklich angesehen werden können, wenn sie nur begrenzte Aufmerksamkeit erfordern und mit einem Blick erfasst werden können. Tabelle 23 verdeutlicht zusammenfassend die Befundlage.

Tabelle 23: Zusammenfassende Befundlage Studie I

Hypothese	Variable	Befundlage
1	Usability	Die Vermittlung von ZF verschlechtert die Beurteilung verglichen zu den anderen Bedingungen. Die zusätzliche Darbietung von IF kann diesen Effekt abmildern.
6	Beanspruchung	
2	Transparenz	
4	Zufriedenheit	
8	Gestaltung	
3	Effektivität	Die Vermittlung von IF erhöht die Anzahl der Fehlbedienungen und damit die Dialogdauer verglichen zu den anderen Bedingungen. Die zusätzliche Darbietung von ZF kann diesen Effekt abmildern.
5	Effizienz	
7	Systemkompetenz	Weder IF, noch ZF haben einen Effekt auf die Memorierungsleistung.
9	Blicke	IF führt zu einem Anstieg der Blickabwendung.
10	SDLP	Weder IF, noch ZF haben einen Effekt auf die Spurhaltekompentenz.

6.1.4.1 Diskussion Umsetzung

Die Befundlage weist darauf hin, dass die Versuchspersonen das **Zustandsfeedback** als nicht zufriedenstellend erachteten. Grundlegend kann dies darauf zurückzuführen sein, dass die vermittelten Informationen der Informativitätsmaxime (siehe Kap. 3.1.2.1) nicht genügten und eher zu viele redundante Informationen darboten. Eine Weiterentwicklung des Zustandsfeedbacks könnte somit darin bestehen, die Rückmeldungen durch Handlungsempfehlungen aufzuwerten. Für zukünftige Systeme wäre denkbar, dass die prozessbegleitenden Zustände durch Meta-Informationen aufgewertet würden, die die Wahrscheinlichkeit einer Fehlerkennung reduzierten. So könnten innerhalb des Receiving-Zustands Informationen über zu laute, zu leise oder überbetonte Spracheingaben bereits dargeboten werden, während der Nutzer noch spricht, und ihm damit Anhaltspunkte zur effizienten Dialogführung geben.

Die Befundlage spricht ebenso dafür, dass das ZF aufgrund der geringen Größe und der schlechten Lesbarkeit bewusst ausgeblendet wurde, da es zu einer zu hohen Beanspruchung führte. Die Blickabwendung, Memorierungsleistung und Erlernbarkeitsbeurteilung in den Zustandsfeedbackgruppen stützen diese Annahme. Die von Clark und Brennan (1991) geforderte Sichtbarkeit der Gesprächspartner konnte die vorliegende Umsetzung somit nicht herstellen und damit auch in dieser Hinsicht den Grounding-Prozess nicht erleichtern.

Bei der geringen Präsenz der Zustände kann ebenfalls nicht davon ausgegangen werden, dass die Warte- oder Gesprächsinitiierungskosten (ebd.) reduziert werden konnten. Da die Systemzustände gegebenenfalls wenig Beachtung fanden, konnten auch die Fehlbedienungen nicht weiter vermieden und die erwartete Reduktion der Fehlerkosten (ebd.) nicht angenommen werden. Es ist davon auszugehen, dass bei besserer gestalterischer Umsetzung des Feedbacks sich auch die Effekte des Zustandsfeedbacks auf Effektivität zeigen lassen. Denn in Kombination mit dem (größeren) Inhaltsfeedback konnten vielversprechende mildernde Interaktionenbewertungen aufgezeigt werden.

Es lässt sich ein äußerst negativer Effekt von zu kleinen, abstrakten Darstellungen berichten. Neben den globalen Usablity-Skalen, wirkte sich die schlecht lesbare Zustandsanzeige auch negativ auf das Belastungsempfinden der Versuchspersonen aus und führte somit zu einer verstärkten Unzufriedenheit. Die Größe der Visualisierungen sollte in folgenden Untersuchungen umfassende Berücksichtigung finden.

Abseits der grafischen Modalität, sollten ebenso alternative Umsetzungen hinsichtlich der mentalen Belastung vergleichend getestet werden. So könnte beispielsweise das System von Hirasawa et al. (1999), welches ausschließlich dialogstrukturierende Signale als Backchannel produzierte als Alternative zur visuellen Dialogergänzung getestet werden.

Das **Inhaltsfeedback** zeigte zwar keine offensichtlichen gestalterischen Mängel, provozierte jedoch durch eine Timing-Schwäche häufig zu frühe Spracheingaben. Dadurch folgten Einbußen in der Effektivität und Effizienz des Systems und es kann davon ausgegangen werden, dass die positiven Effekte des Inhaltsfeedbacks dadurch überlagert wurden. Auch hier sollten die Vorteile einer textuellen Rückmeldung mit einer besseren zeitlichen Umsetzung des Feedbacks erneut evaluiert werden.

Weiterhin lassen sich Kontexte, wie beispielsweise die Interaktion mit einem Diktiererkenner, antizipieren, in denen Nutzer von einer textuellen Darbietung mehr profitieren und diese auch besser bewerten würden (vgl. Curin et al., 2011).

Da die akustische Rückmeldung des Systems bereits alle relevanten Informationen enthielt, wurde die Visualisierung der Dialoginhalte häufig als überflüssig und nur bedingt hilfreich erachtet. So wurde insbesondere die Darstellung der eigenen Spracheingabe als wenig hilfreich angesehen. Insgesamt kann dies auf die limitierte Handlungsfähigkeit des Nutzers durch die begrenzte **Dialogfunktionalität** des Systems zurückgeführt werden. In der vorliegenden Umsetzung konnten Fehlerkennungen durch die Anzeige der Dialoginhalte durch den Nutzer zwar früher erkannt werden, sofortige Korrekturen waren allerdings nicht möglich. Damit bot das Inhaltsfeedback zwar eine transparente Fehlerrückmeldung, jedoch keine Möglichkeit zur Reparatur des Dialogs. Konkret bedeutet dies, dass der ergänzende Charakter der Visualisierung nur dann zur Minimierung der Reparaturkosten (Clark & Brennan, 1991) beitragen kann, wenn eine Funktionserweiterung des Systems stattfindet. Die Umsetzung eines realen Barge-Ins wäre an dieser Stelle anzuraten, um dem Nutzer die sofortige Gelegenheit zur Korrektur einräumen zu können, wenn er ein Missverständnis bemerkt. Denkbar wäre ebenfalls die Implementation einer Korrekturmöglichkeit innerhalb der Sprechblasen über die grafisch-haptischen Bedienschnittstellen, um im Fehlerfall das Ausweichen auf alternative Eingabemethoden zu erleichtern.

Eine weitere Aufwertung der Funktionalität des Inhaltsfeedbacks könnte auch durch das Hervorheben von feldfüllenden Schlagworten, die für das System zur Aktionsausführung benötigt werden, geschehen (siehe Abbildung 37). Sie könnten dem Nutzer als Memorierungshilfe dienen und mit dem speak-what-you-see-Konzept verbunden werden. Ergänzend könnten auch jene Worte markiert werden, die schlecht oder gar nicht verstanden wurden, um die Systemgrenzen zu verdeutlichen und Reparaturmöglichkeiten anzubieten (vgl. Clark & Brennan, 1991).

Abbildung 37: Umsetzungsbeispiel für die Hervorhebung feldfüllender Informationen

Da es zu einer höheren Blickabwendung und visuellen Beanspruchung bei der Darbietung der Dialoginhalte kam, kann ebenso nicht von einer Reduktion der Rezeptionskosten nach Clark und Brennan (1991) gesprochen werden. Nur in Kombination mit dem Zustandsfeedback gaben die Probanden an, immer darüber informiert gewesen zu sein, was das System verstanden hat. Durch die Timing-Schwäche konnte die Darbietung der akustischen und visuellen Systemausgabe nicht simultan geschehen. Damit konnte die Rezeption der akustischen Sprachausgabe nicht

durch die textuelle Darbietung unterstützt werden. Eine sequentielle Ausgabe schien in diesem Kontext nicht erfolgsversprechend zu sein.

Durch die Umsetzung des persistenten Inhaltsfeedback hätte die Möglichkeit bestanden, sich die Systemausgabe zu einem späteren Zeitpunkt noch einmal anzuschauen. Diese Auflösung der sprachlichen Vergänglichkeit (ebd.) erleichtert das Grounding mit dem System vor allem dann, wenn der Nutzer der akustischen Ausgabe aufgrund eines komplexen Fahrmanövers nicht folgen konnte. Dieser Fall, in dem der Nutzer stark von dem Inhaltsfeedback profitiert hätte, kam im Rahmen der vorliegenden standardisierten Untersuchung jedoch kaum vor. So kann die Fahraufgabe im Allgemeinen als sehr leicht bezeichnet werden. Es gab in den seltensten Fällen kritische Fahrmanöver, die die Versuchspersonen von der Interaktion mit dem SDS ablenkten. Häufig konnte auch beobachtet werden, dass Probanden erst dann den Dialog starteten, wenn sie keine anspruchsvollen Verkehrssituationen mehr antizipierten.

Die **kombinierte Darbietung** beider Feedbackarten konnte dagegen in vielen Fällen positive Verbesserungen verglichen zu der alleinigen Anzeige der Systemzustände und Dialoginhalte erzielen (siehe Tabelle 23). Eine kombinierte und damit größere Darstellung konnte somit zwar nicht die erwartete positive Abweichung von der Kontrollgruppe erzielen, aber die negative Abweichung durch die alleinige Präsentation ausgleichen.

Zur Bewertung der untersuchten Feedbackarten wurde keine Anbindung des Sprachdialogsystems an die reale grafische Anzeige des Adressbuchs vorgenommen. Eine Integration und Übertragung des hier postulierten Rückmeldekonzepts auf weitere Kontexte ist inhaltlich vorstellbar, sollte allerdings hinsichtlich des Ablenkungspotentials innerhalb des Gesamtkonzepts erneut adressiert werden. Auch wenn die Ablenkung der Anzeige bisher als moderat einzustufen wäre, muss im Kontext eines multimodalen Bedienkonzeptes permanent davon ausgegangen werden, dass weitere Menüinhalte visuell und häufig auch textuell präsentiert werden. Des Weiteren könnte man bei einer textuellen Inhaltsanzeige eine Interferenz mit dem speak-what-you-see-Konzept, welches dem Nutzer die kontextabhängigen Kommandos vorschlägt, erwarten. Eine erneute Absicherung des integrierten Visualisierungskonzepts sei angeraten.

6.1.4.2 Diskussion Methode

Der SUMI-Fragebogen sollte primär dazu verhelfen, die Eindrücke und Gefühle der Nutzer bei der Systemnutzung detaillierter abzubilden als der SUS. So konnten die globalen Gebrauchstauglichkeitsmängel des Zustandsfeedbacks auf spezifische Unterkategorien der Usability abgebildet werden. Einschränkend muss Erwähnung finden, dass der SUMI keinen Bezug zu international gültigen Normierungen besitzt und die Auswertung sich eher unhandlich gestaltete.

Obwohl Kirakowski (1994) für den SUMI ein Intervallskaleniveau annimmt, muss dies bei der gegebenen Skalenbeschriftung hinterfragt werden. So fließen in die mittlere Kategorie ("weiß nicht") zwei Aspekte ein. Einerseits konnte die Versuchsperson diese Kategorie nutzen, wenn sie ihre Neutralität zum Ausdruck bringen möchte und andererseits kann sie damit auch zum Ausdruck bringen, wenn sie eine Aussage nicht beurteilt kann, da eine spezifische Situation mit dem System nicht erlebt wurde. Die mittlere Kategorie kann demnach nicht als neutraler Nullpunkt und somit für den SUMI nur das Ordinalskalenniveau angenommen werden.[24]

Durch die große Anzahl an Items verlängerte der SUMI den vorliegenden Fragebogen beachtlich, so dass hier auch ein Ermüdungseffekt der Probanden nicht ausgeschlossen werden kann. Da er eine hohe Übereinstimmung mit dem SUS aufwies, wird die Anwendung für die zukünftigen Studien nicht vorgesehen.

Bei der Überprüfung der Subskalen des SUMI zeigte sich, dass das Konstrukt Effizienz nicht als eindimensional betrachtet werden kann. So lieferte eine angeschlossene Faktorenanalyse (Anhang 12.2.5) eine dreifaktorielle Lösung, bei der unter anderem ein Faktor entsteht, den man deutlich als zeitbezogene Effizienz bezeichnen könnte (System reagiert zu langsam, Geschwindigkeit des Systems ist hoch genug, sparsame Eingabebefehle möglich). Die anderen beiden Faktoren decken eher die Handhabbarkeit des Systems und die Transparenz der Informationsvermittlung ab.

Darüber hinaus geschah die subjektive Beurteilung der Effektivität über ein Item, welches hierfür nicht primär konzipiert wurde (*Das System versteht mich gelegentlich unerwartet falsch.*). Innerhalb der Formulierung des Items fand sich noch eine zeitlich Einschränkung (*gelegentlich*), welche die Zustimmung künstlich erhöht. In zukünftigen Untersuchungen sollten zur Erfassung der subjektiven Effektivitäts- und Effizienzbeurteilung spezifischere Items generiert werden, die direkt die Zuverlässigkeit bzw. Angemessenheit der Interaktionslänge erfragen.

Auch die selbstkonstruierte Skala **Systemtransparenz** konnte keine hohe interne Konsistenz aufweisen (Anhang 12.2.5). Eine Faktorenanalyse zeigte schließlich die Verteilung der Items auf drei Faktoren. Interessant ist, dass sich ein Faktor auf die Zustände des System (*Ich wusste immer was das System tut / ich tun kann*) und ein Faktor auf die Systemkompetenz (*Ich wusste immer was das System gesagt hat / Ich habe ein Gefühl dafür, was das System kann.*) bezog.

[24] Angeschlossene Analysen unter Verwendung eines nicht-parametrischen Tests (Kruskal Wallis Varianzanalyse) zeigten in den meisten Fällen keine signifikanten Gruppenunterschiede.

Den dritten Faktor bildete das Item *Ich wusste immer was das System verstanden hat.* Es zeigte sich also, dass sich die Skala deutlich in die Faktoren Zustandsvermittlung und Inhaltsvermittlung trennen lässt, während bei letzterem ein Unterschied zwischen sprachlicher Kompetenz und Erkennerleistung des Systems vorzuliegen scheint.

Die **Künstlichkeit der Laborsituation** wirkte sich in vielerlei Aspekten auf die untersuchte Mensch-Maschine-Interaktion aus. Durch das Vorsprechen der Kommandos durch den Versuchsleiter wurde die Formulierungsphase für den Nutzer beachtlich erleichtert. So musste er keinen initialen Common Ground aufbauen und antizipieren, was das System verstehen könnte, sondern wiederholte lediglich die Aufforderung des VL. Dabei wurden nicht nur OOV-Fehler drastisch reduziert, sondern auch phonetische Besonderheiten beeinflusst. Durch Alignmentprozesse (vgl. Kap. 3.2.3) hat sich die Versuchsperson stark an die Aussprache des VLs angepasst. Die Wahrscheinlichkeit für anfänglich hyperartikulierte Äußerungen kann dadurch als reduziert betrachtet und Fehlerkennungen unterschätzt werden.

Es kann weiterhin davon ausgegangen werden, dass zu frühe Spracheingaben und Substitutionsfehler auch durch die der Versuchsperson **fremden Namen** im Adressbuch begünstigt wurden. So kam es vor, dass die Probanden den VL schlecht verstanden und somit bereits mit einem falschen Namen in den Dialog starteten. Einige Probanden schienen mit der Memorierung des Kommandos bzw. des Namens Probleme zu haben.

Die Artifizialität der Laborsituation führte auch dazu, dass die Probanden auf namentliche Ersetzungen (Substitutionsfehler des Systems) mit größerer Akzeptanz reagierten als unter normalen Umständen. Während ein falscher Telefonanruf in der Realität peinlich ist, war er in dem gegebenen Untersuchungskontext unproblematisch und ohne Konsequenzen.

In der vorliegenden Untersuchung konnte keine **Gleichverteilung der Geschlechter** erreicht werden. So sind Frauen stark unterrepräsentiert, wenn auch gleichverteilt auf die Untersuchungsbedingungen. Auch wenn theoretisch kein Unterschied in der Systembeurteilung erwartet wird, so sollte in den folgenden Untersuchungen eine Gleichverteilung angestrebt werden.

6.1.5 Zusammenfassung

Obgleich bestätigt werden konnte, dass ergänzende Visualisierungen eines SDS im Fahrkontext unter den gegebenen Umsetzungen als unbedenklich für die Fahraufgabe angesehen werden können und sich Probanden der KG diese wünschten, so konnte nicht gezeigt werden, dass die Implementation einer ergänzende Visualisierung eine erfolgsversprechende Stellgröße der Dialoggestaltung ist. Eine Erleichterung der Grounding-Prozesse oder die Reduktion der nutzerseitigen Aufwandskosten konnte nicht nachgewiesen werden.

Dabei muss in Betracht gezogen werden, dass die vorliegende Untersuchung unter einem Gestaltungseffekt leidet. Die adressierten Stellschrauben zur Erleichterung des Grounding-Prozesses (Sichtbarkeit, Vergänglichkeitsvermeidung) und zur Reduktion der Prozesskosten (Fehler-, Reparatur-, Rezeptionskosten) konnten durch die gestalterischen Mankos nicht angenommen werden. Die ursprünglichen Forschungsfragen, ob die Vermittlung visueller Zustands- oder Inhaltsinformationen den Mensch-Maschine-Dialog verbessern kann, können vor diesem Hintergrund nur bedingt beantwortet werden. Die korrespondierenden Empfängerhandlungen sind so zeit- und situationsabhängig, dass eine Übertragung auf maschinelle Dialogpartner nur bei der Beachtung sensibler Gestaltungsparameter erfolgreich sein kann. Spätere Untersuchungen sollen eine derartige Verzerrung durch die Umsetzung ausschließen.

Insgesamt konnten jedoch Hinweise darüber abgeleitet werden, dass visuelle Rückmeldungen das Dialogverhalten von Nutzern aktiv beeinflussen können. Unterstützt wurde die Mensch-Dialogsystem-Interaktion vor allem dann, wenn Zustands- und Inhaltsfeedback kombiniert dargeboten wurden. In Anlehnung an die zwei Dialogspuren nach Clark (1996) kann geschlussfolgert werden, dass auch in der Mensch-Maschine-Kommunikation sowohl den Zustandsindikatoren als auch den Inhalten, eine zentrale Bedeutung zukommt.

Es lassen sich die folgenden Gestaltungsempfehlungen ableiten:

- Die Größe und das Timing der visuellen Anzeigen sind sensible Gestaltungsfaktoren, die nicht verletzt werden dürfen.
- Rückmeldekonzepte sollten sowohl Zustands- als auch Inhaltsfeedback beinhalten.
- Die akustische und visuelle Darbietung der Systeminhalte sollte parallel-redundant geschehen.
- Bei gegebenem Inhaltsfeedback sollten dem Nutzer Möglichkeiten zur direkten Korrektur eingeräumt werden (z. B. Barge-In).

Nachdem die Ergebnisse nun hinsichtlich der ursprünglichen Hypothesen, dem Untersuchungskontext und der Umsetzung des Systems diskutiert wurden, sollen im folgenden Absatz Schlussfolgerungen für die folgenden Untersuchungen abgeleitet werden.

Einflüsse auf die folgende Studie

Neben einer breiten Befundlage (Cameron, 2000; Brumby, 2011; Wechsung et al., 2010) zeigte sich die **Effizienz** auch in der vorangegangenen Untersuchung als eine zentrale Determinante der Gebrauchstauglichkeit. Diese Anregung wurde in der Studie II aufgenommen, in der die Effizienz als Stellschraube der Dialoggestaltung getestet werden soll. Durch die Flexibilisierung der

Interpretationsstrategie, soll nun aktiv in die Dialogführung eingegriffen und diese damit nicht nur begleitend unterstützt werden.

In den folgenden Studien soll sich die **Effektivität** der SDS auf einem konstant hohen Niveau einpendeln. Somit wurde nicht nur die Erkennerleistung, sondern auch der Umgang mit fehlenden Eingaben und/oder TTE-Fehlern verbessert. Die Reduktion der Telefonbucheinträge und die Konzentration auf weniger Funktionen des Adressbuchs sollen robustere Dialoge ermöglichen. Diese Anpassungen waren notwendig, um die Änderungen am Dialogdesign evaluieren zu können, da sie bei der unterdurchschnittlichen Erkennerleistung der ersten Erhebung kaum Beachtung finden konnten.

In der nachfolgenden Studie soll ebenfalls eine ergänzende **Visualisierung** zum Einsatz kommen. Aufbauend auf den vorliegenden Erkenntnissen, wurde eine reduzierte Rückmeldung des kombinierten Zustands- und Inhaltsfeedbacks umgesetzt. Um die Blickabwendung bei der textuellen Darbietung der Dialoginhalte weiter zu reduzieren, wurde sich auf eine Anzeige der zentralen Schlüsselbegriffe beschränkt. Dadurch wird auch eine kombinierte Anzeige der erkannten Nutzereingabe und der Systemausgabe möglich, die sich bei dieser feldbasierten Anzeige nicht mehr unterscheiden dürften. Da keine separate Anzeige der Systemausgabe erfolgte, erübrigt sich das Timing-Problem aus Studie I.

Das Zustandsfeedback wird nun auf den Ready- und Speaking-Zustand beschränkt und sehr viel größer dargeboten. Die bisher vorgenommene künstliche Verlängerung des Processing-Zustands findet nicht mehr statt. Eine Anzeige des Verarbeitungszustands wird somit für die adressierten Kontexte überflüssig. Weiterhin wurde ein sehr viel deutlicherer Aufbau des Ready-Zustands umgesetzt um TTE-Fehler bei der ersten Eingabe zu vermeiden. Der Übergang von dem Druck des PTT-Knopfes bis zur Empfangsbereitschaft des Systems wird nun durch eine großflächigere Animation begleitet.

Da die visuelle Überbrückung der systemseitig-bedingten Wartephase mittels Zustandsfeedback nur mäßig erfolgreich war, soll nun ein aktiver Eingriff in das Dialogdesign erfolgen. Studie II adressiert mit der Umsetzung des System Grounding Criterions primär die Effizienz des Dialogs und wird im Folgenden ausführlich dokumentiert.

6.2 Studie II: System Grounding Criterion

Im Rahmen der zweiten Untersuchung sollte ein System realisiert werden, das nur noch dann eine Bestätigungsfrage äußert, wenn es sich bei der Erkennung einer Nutzereingabe unsicher ist. Die Stellgröße der Erkennerkonfidenz beeinflusst bei jedem Dialogschritt, ob das System eine defensive Interpretation des Ergebnisses (und den Nutzer um eine explizite Verifikation bittet) oder eine offensive Interpretation zeigt und im Dialog fortfährt. Jenes dynamische System soll in der folgenden empirischen Untersuchung gegen ein System mit konstant defensiver Interpretationsstrategie (im Folgenden: statisches System) verglichen werden. Es leiten sich die folgenden Fragestellungen ab.

6.2.1 Hypothesen

Durch die Umsetzung des Prinzips des Least Collaborative Effort (Clark & Wilkes-Gibbs, 1986) werden der Grounding-Prozess erleichtert und die nutzerseitigen Aufwandskosten minimiert (Clark & Brennan, 1991). Durch die Reduktion der Dialogschritte verringern sich nicht nur die Rezeptions-, sondern auch die Wartekosten für den Nutzer. Primär führt ein liberales System Grounding Criterion zur Vermeidung überflüssiger Bestätigungsanfragen und damit zu einer Effizienzerhöhung:

1. **Eine dynamische Dialoggestaltung erhöht die objektive und subjektiv empfundene Effizienz der Mensch-Maschine-Interaktion.**

Es wird weiterhin erwartet, dass durch eine sinnvolle Anpassung des Systemverhaltens an die Dialogsituation eine Aufwandsreduktion bei dem Nutzer erreicht werden kann, die nicht nur zeitlicher Natur ist. Bedingt durch gewisse Fehlerwahrscheinlichkeiten bei der Interaktion mit einem SDS reduziert sich mit der Anzahl der Turn Takes auch die Möglichkeit für Bedienfehler oder Fehlerkennungen. Durch die reduzierte Anzahl an Sprecherwechseln sinken somit die Gelegenheiten für Fehler jeglicher Art, wovon die Effektivität des Dialogs profitiert:

2. **Eine dynamische Dialoggestaltung erhöht die Effektivität der Mensch-Maschine-Interaktion.**

Von der gegebenen Effektivität und Effizienz des Dialogs wird auch die Nutzerzufriedenheit profitieren, da die flexible Anpassung der Interpretationsstrategie dem kollaborativen Charakter der Kommunikation gerecht wird und das System Verantwortung für die Verringerung des geteilten Aufwands übernimmt:

3. **Eine dynamische Dialoggestaltung verbessert die Systembewertung und erhöht die nutzerseitige Zufriedenheit mit dem System.**

Aus diesen drei Annahmen ergibt sich die Hypothese, dass ein dynamisches System die Gebrauchstauglichkeits-Kriterien nach DIN EN ISO 9241-11 (1998) besser erfüllt, als ein System mit statischen Rückfrageverhalten:

4. **Eine dynamische Dialoggestaltung verbessert die allgemeine Gebrauchstauglichkeitsbewertung des Systems.**

In diesem Zusammenhang wird angenommen, dass die dynamische Dialoggestaltung die verhaltensbasierte Bereitschaft Sprachbedienung zu benutzen erhöhen sollte:

5. **Die Nutzungsbereitschaft von Sprachdialogsystemen wird durch beide Systeme erhöht, während das dynamische System einen zusätzlichen Anstieg gegenüber dem statischen System zeigt.**

Dennoch wird angenommen, dass Nutzer eine gewisse Gewöhnungsphase benötigen und erst nach einigen Systeminteraktionen ausreichend Vertrauen zum System aufgebaut haben, um das dynamische Dialogverhalten zu akzeptieren. Erst nach einer gewissen Lernphase werden Nutzer das kollaborative System besser bewerten als das statische System.

6. **Nach einer Gewöhnungsphase wird das dynamische System hinsichtlich seiner Konsistenz, Transparenz, Intuitivität und Verständlichkeit besser bewertet.**

Adressiert wird auch die globale Frage nach der obligatorischen Systemkonsistenz. An dieser Stelle wird die Hypothese vertreten, dass eine Systeminkonsistenz durch eine dynamische Anpassung an die Systemsicherheit keine negativen Auswirkungen auf die Systembewertung zeigt, da sie den natürlichen Kommunikationsstrategien entspricht. Es wird davon ausgegangen, dass die Effizienz bei der Dialoggestaltung über der Konsistenz stehen sollte.

7. **Eine dynamische Anpassung des Rückfrageverhaltens hat keine negativen Auswirkungen auf die Konsistenzbewertung des Systems.**

Es soll weiterführend untersucht werden, ob unterstützende visuelle Anzeigen der Systemsicherheit zu einer Verbesserung der Systemtransparenz beitragen, indem sie die Systemdynamik begründen und damit das Konsistenzdilemma abschwächen.

8. **Die visuellen Anzeigen der Systemsicherheit erhöhen die Systemtransparenz des dynamischen Sprachdialogsystems.**

In diesem Zusammenhang soll geprüft werden, ob die Konfidenzvisualisierungen vorherige Erfahrungen und Schemata der Nutzer ansprechen und mentale Modelle auch von jenen Probanden erstellt werden können, die im Versuch mit anderen Anzeigen konfrontiert wurden:

9. **Gruppen ohne Anzeigen der Systemsicherheit zeigen ähnliche Assoziationen wie die Gruppen mit Anzeigen.**

Dabei wird angenommen, dass weder die dynamische Dialogstruktur, noch die ergänzende Konfidenzvisualisierung die Ablenkung von der Fahraufgabe erhöhen oder die Blickabwendung begünstigen:

10. **Die dynamische Dialogstruktur und die ergänzende Konfidenzvisualisierung zeigen keine negativen Effekte auf die objektiven Fahrparameter**
11. **oder auf die prozentuale Anzahl der Blickabwendungen.**

Explorativ soll anhand korrelativer Zusammenhänge beurteilt werden, ob **Hyperartikulation** als objektive Messung der Zufriedenheit geeignet ist. Das folgende Kapitel soll Einblicke in den Untersuchungskontext der zweiten Studie geben.

6.2.2 Methode

Die Untersuchung fand, wie die erste Untersuchung, im feststehenden Fahrsimulator der Volkswagen Konzernforschung statt. Die Fahraufgabe wurde identisch zur ersten Studie umgesetzt. Deren Beschreibung und die des Simulators können Kapitel 6.1.2 entnommen werden.

6.2.2.1 Sprachdialogsystem

In Anlehnung an die Überlegungen von Brennan und Hulteen (1995), sollte ein Ziel dieser Arbeit sein, ein optimales System Grounding Criterion (SGC) für das vorliegende Sprachdialogsystem zu definieren. Insgesamt wird davon ausgegangen, dass es unmöglich ist, eine perfekte initiale Dialogkonfiguration des SGCs zu finden, die für alle Nutzer in allen Situationen angemessen ist. Das folgende Kapitel beschreibt daher, wie eine system-initiierte Anpassung der Dialogstrategie in Abhängigkeit der Gegebenheiten innerhalb eines Dialogs realisiert wurde. Der Untersuchung lag erneut das in Kapitel 6.1.2.2 beschriebene Dialogsystem zugrunde.

6.2.2.1.1 Umsetzung eines flexiblen SGC

In Anlehnung an zwischenmenschliche Dialoge soll je nach Empfängerzustand über den weiteren Fortgang des Dialogs entschieden werden. Dabei erscheint die **Erkennerkonfidenz** als ein viel-

versprechender Schätzer der Systemsicherheit bei adäquatem technischem Aufwand und dient in der vorliegenden Arbeit der Definition des SGCs.

Nach der Spezifikation von Benutzermodellen (BM) von Kass und Finin (1988) handelt es sich bei der Anpassung des SGC an die Erkennerkonfidenz somit um ein generisches und dynamisches Kurzzeit-BM, das alle Nutzer als eine homogene Grundgesamtheit betrachtet und keine individuellen Anpassungen auf Basis einer Dialoghistorie vornimmt. Zur system-initiierten Anpassung werden die Informationen nur während einer Spracheingabe erstellt und anschließend wieder verworfen.

Wird das System Grounding Criterion nun an die Sicherheit des Spracherkenners bei der letzten Eingabe geknüpft, so beeinflusst diese Konfidenz bei jedem Dialogschritt, welche systemseitigen Empfängerzustände nach dem adaptierten Contribution Modell (Kapitel 4.2.1) zurückgemeldet werden. Je nach Höhe der Konfidenz wird adaptiv entschieden, ob das System eine explizite Bestätigungsaufforderung an dieser Stelle äußert oder im Dialog fortfährt. Statt ermüdender expliziter Bestätigungsanfragen wird somit bei einer hohen Erkennerkonfidenz die gesamte Handlungsplanung (z. B. das Starten eines Telefonanrufs) direkt angekündigt (vgl. Brennan & Hulteen, 1995). Systemnachfragen und Bestätigungsaufforderungen finden im Rahmen dieser Umsetzung nur noch dann statt, wenn ein erhöhter nutzerseitiger Aufwand gerechtfertigt werden kann. Dies ist insbesondere dann der Fall, wenn eine bestimmte Konfidenzschwelle unterschritten wird und damit ein Missverständnis wahrscheinlich wird. Damit sollte dem Nutzer bei unsicherer Erkennung die Möglichkeit zur Korrektur durch eine Bestätigungsanfrage gegeben werden. So werden überflüssige Interaktionsschritte im Sinne des Least Collaborative Effort Ansatzes gespart und auch die Maxime der Quantität, Relevanz und der Art & Weise berücksichtigt (vgl. Kap. 3.1.2.1).

Zwei **Konfidenz-Schwellen** sollen demnach zur Umsetzung des SGC definiert werden. Einerseits kann unter einer gewissen Konfidenz keine Interpretation der Nutzereingabe stattfinden. Auf diesen Zustand, der durch ein *no match* oder *no input* (siehe Kapitel 2.2.3) hervorgerufen werden kann, folgt eine erneute Eingabeaufforderung (*Wie bitte?*). Diese Schwelle1, die auch als Wortfehlerratenschwelle bezeichnet wird, ist gängiger Bestandteil der meisten gegenwärtigen SDS.

Andererseits soll zur Flexibilisierung des Rückfrageverhaltens die Einführung einer zweiten Konfidenzschwelle (Schwelle2) erfolgen. Diese definiert, wie sicher ein System sich mit der Interpretation der Nutzereingabe ist und liegt damit immer über der ersten Schwelle. Im Bereich zwischen Schwelle1 und Schwelle2 wird das System aufgrund seiner Unsicherheit mit einer Bestätigungsanfrage reagieren und nach einer expliziten Bestätigung (*Möchten Sie Helga Otto anrufen?*) verlangen. Auf alle Nutzereingaben, deren Konfidenz die Schwelle2 übertreffen, reagiert das System direkt mit der Ankündigung der Handlungsplanung (*Helga Otto. Es wird gewählt.*). Durch die Ein-

führung dieses zweiten Konfidenzschwellenwertes kann ein flexibles System Grounding Criterions erreicht und somit eine hohe Dynamik im Dialogverhalten realisiert werden.

Es sollte bei der Umsetzung allerdings Beachtung finden, dass bei einer hohen Erkenner-seitigen Konfidenz nicht zwangsläufig eine korrekte Interpretation des Nutzerinputs vorliegt (Litman & Pan, 2002). Auch niedrige Konfidenzwerte deuten nicht notwendigerweise auf eine falsche Erkennung hin. In Anlehnung an die Signalentdeckungstheorie (Green & Swets, 1966) bedeutet dies, dass auch falsche Inhalte mit hoher Konfidenz (engl. *false positives*) verstanden oder richtige Inhalte mit geringer Konfidenz (engl. *misses*) zurückgewiesen werden können (siehe Abbildung 38).

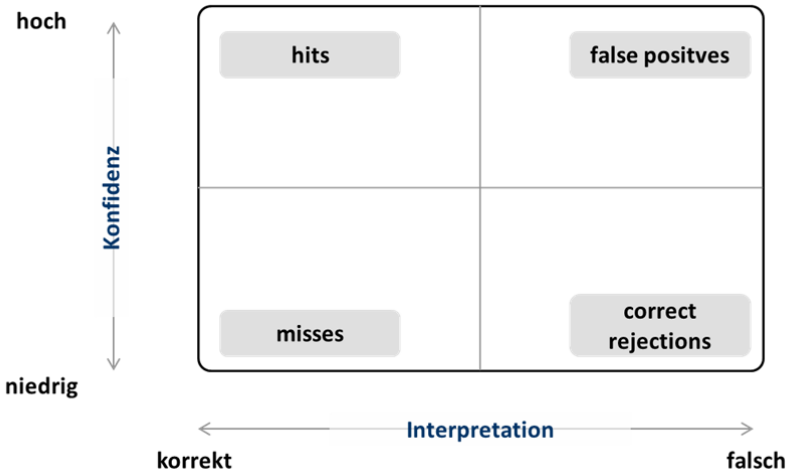

Abbildung 38: Konfidenz-Verständnis-Kontinuum

Im Rahmen einer **Vorstudie** sollte deshalb ein repräsentativer Konfidenzschwellenwert für die Schwelle2 identifiziert werden, der einerseits die Dialoge signifikant verkürzt, aber andererseits keine false positives begünstigt. Jene Substitutionsfehler würden eine Handlungsplanung für eine falsche Aktion ohne Rückversicherung ankündigen und die Effektivität des Systems gefährden. Die gesuchte Konfidenzschwelle sollte somit eine optimale Balance zwischen Effizienz und Effektivität bieten. Auf Basis von Konfidenzanalysen vorangegangener Studien wurden drei Schwellenwerte identifiziert, die in jenem Prä-Test sinnvoll gegeneinander getestet werden sollten. Das konservativste SGC lag bei 70%. Eine mittlere Schwelle lag im Durchschnittsbereich der zuvor aufgezeichneten Konfidenzen des SDS bei 55% und das liberalste SGC lag bei 40%. In der liberalen SGC-Bedingung reichte eine Konfidenz größer als 40% aus, um eine explizite Bestätigungsanfrage zu vermeiden und eine Handlungsausführung direkt einzuleiten, während dafür in der konservativen Bedingung 70% benötigt wurden.

Auf Basis der Vorstudienergebnisse, deren vollständige Beschreibung Anhang 12.3 entnommen werden kann, fand die liberalste Konfidenzschwelle (40%) Anwendung. Die Analysen zeigten, dass die liberalste Schwellendefinition zu der größten Effizienz- und Effektivitätssteigerung führte, ohne die Anzahl der false positives zu erhöhen. Das liberalste SGC zeigte darüber hinaus die besten subjektiven Bewertungen. Auch die durchschnittliche Konfidenz der Kommandos von 45.9% (*SD*= 7.07) spricht für den hohen Differenzierungsgrad dieser Schwelle.

Das vorliegende Sprachdialogsystem, dessen ursprüngliche Konfiguration eine defensive Interpretationsstrategie aufwies, kann nun adaptiv, je nach Sicherheit, bei der Interpretation der Nutzereingabe eine defensive oder offensive Dialogstrategie zeigen und sich ein Kommando implizit oder explizit bestätigen lassen. Mit dieser Umsetzung lassen sich nun folgenden Dialogbedingungen realisieren.

6.2.2.1.2 Dialogbedingung

Zur Evaluation der Effekte dynamischer Dialoggestaltung wurden zwei Varianten, basierend auf verschiedenen SGCs, konzipiert.

Statisches Dialogverhalten, wie es aktuelle Sprachdialogsysteme zeigen, wurde durch eine konservative Unsicherheitsschwelle von 100% erreicht[25]. In dieser Bedingung fragte das System für jede Aktion nach einer Bestätigung und zeigte damit eine defensive Interpretation der Erkennerergebnisse. Erst nach der expliziten Verifikation des Nutzers wurde die Handlungsausführung angekündigt.

In der **dynamischen Dialogbedingung** wurde eine liberale Konfidenzsschwelle in Anlehnung an die Ergebnisse der Vorstudie definiert. Sofern die Konfidenz des Erkenners in dieser Bedingung 40% überschritt, wurde sofort die Handlungsausführung kommuniziert. Unterschritt die Konfidenz die geforderten 40% so reagierte das System mit einer Bestätigungsanfrage.

In den Tabelle 24 und Tabelle 25 soll das Dialogverhalten in Abhängigkeit der Konfidenz und der Dialogbedingung an einem Beispiel dargestellt werden.

---

[25] Das statische Systemverhalten stellt auch das Ausgangsverhalten des Sprachdialogsystems in Kapitel 6.1.2.2 dar.

Tabelle 24: Statische oder dynamische Dialogbedingung (Konfidenz< 40%)

Nutzer	Ich möchte Helga Otto mobil anrufen.
System	Möchten Sie Helga Otto mobil anrufen?
Nutzer	Ja.
System	Helga Otto. Es wird gewählt.

Tabelle 25: Dynamisches Dialogverhalten (Konfidenz> 40%)

Nutzer	Ich möchte Helga Otto mobil anrufen.
System	Helga Otto. Es wird gewählt.

6.2.2.1.3 Feedbackbedingung

Zusätzlich zur Sprachausgabe lieferte das System auch visuelle Rückmeldungen an die Nutzer. Sowohl das Inhalts- als auch Zustandsfeedback (bekannt aus der ersten Erhebung) wurde erneut im Display des Kombiinstruments angezeigt.

Die textuelle Anzeige des **Inhaltsfeedbacks**, die nur feldfüllende Informationseinheiten und keine Füllworte enthielt, war dabei für alle Feedbackbedingungen gleich. Das erweiterte **Zustandsfeedback**, das Informationen über den Systemzustand bzw. die verfügbaren Handlungsmöglichkeiten enthielt, wurde teilweise variiert.

Abbildung 39: Beispiele für die Umsetzung des Zustands- und Sicherheitsfeedbacks

Während ein graues Mikrofon in beiden Bedingungen die Aufnahmebereitschaft für die Nutzereingabe (Ready) anzeigte, variierte die Farbe der Lautsprecher (Speaking), die erschienen wenn das System sprach. In einer Feedbackbedingung wurde ausschließlich das reine Zustandsfeedback (graues Mikrofon und grauer Lautsprecher) vermittelt. In der zweiten Feedbackbedingung

wurde zusätzlich zu diesem Zustandsfeedback sogenanntes Konfidenz- oder Sicherheitsfeedback vermittelt, in dem die Systemsicherheit durch eine Färbung der Lautsprecher dargestellt wurde.

Da sich die Ampel-Metapher gut mit der Dreistufigkeit des vorliegenden Schwellenkonzeptes deckt und der Farbwechsel peripher gut wahrnehmbar ist wurde sich für diese Art der Visualisierung der systemseitigen Sicherheitslevels entschieden (siehe Abbildung 39). Welche Farbe die Icons in der Konfidenzvisualisierungsgruppe annahmen, war einerseits an die Erkennerkonfidenz geknüpft. In Anlehnung an die Verständnisevidenzen der ersten Studie (vgl. Kap. 6.1.2.2) stand Orange in beiden Dialogbedingungen für den Fall einer Unterschreitung der Schwelle1 („keine Eingabe/ keine Übereinstimmung gefunden“). Andererseits ergaben sich abhängig vom Dialogverhalten leicht unterschiedliche Bedeutungen des gelben und grünen Lautsprecher-Icons.

In der statischen Bedingung signalisierte die Gelbfärbung aufgrund der konservativen Schwelle eine Bestätigungsanfrage. Auf diese folgte in Grün die Ankündigung der Aktionsausführung (siehe Tabelle 26).

Tabelle 26: Typischer Dialogablauf in der statischen Bedingung

Nutzer	Ich möchte Helga Otto mobil anrufen.	Konfidenz 0-100%	
System	Möchten Sie Helga Otto mobil anrufen?		
Nutzer	Ja.	Konfidenz 0-100%	
System	Helga Otto. Es wird gewählt.		

In der dynamischen Bedingung zeigte Gelb hingegen an, dass der Nutzer nicht ausreichend gut oder vollständig verstanden wurde um die Aktion sofort auszuführen (Konfidenz< 40%). Grün signalisierte dagegen in dieser Bedingung, dass der Nutzer hinreichend gut verstanden wurde und die Aktion direkt ausgeführt werden kann (Konfidenz> 40%). Tabelle 27 verdeutlicht den adaptiven Gesprächsverlauf.

Tabelle 27: Möglicher Dialogablauf in der dynamischen Bedingung

Nutzer	Ich möchte Helga Otto mobil anrufen.	Konfidenz< 40%	
System	Möchten Sie Helga Otto mobil anrufen?		
Nutzer	Nein, ich möchte Laura Albrecht mobil anrufen.	Konfidenz> 40%	
System	Laura Albrecht. Es wird gewählt.		

Innerhalb der dynamisch-farbigen Bedingung lehnt sich die Umsetzung in abstrakter Art und Weise an die Empfangsbestätigungen (z. B. Mimik) der zwischenmenschlichen Kommunikation an und soll somit den adaptiven Wechsel des Systems zwischen der offensiven oder defensiven Interpretation der Nutzereingabe begründen.

Ausschließlich in der dynamischen Bedingung vermittelt die Färbung der Icons demnach einen Informationszugewinn, indem sie die Systemsicherheit anzeigt und das dynamische Dialogverhalten erklärt. In der statischen Bedingung dagegen transportiert die farbige Gestaltung der Zustände keine zusätzliche Information, die nicht auch akustisch redundant dargeboten wird.

6.2.2.1.4 Kommandos

Basierend auf Erfahrungswerten der Vorstudie und weiterer vorangegangener Untersuchungen mit dem gleichen Sprachdialogsystem wurden zehn Bedienkommandos aus dem Adressbuchkontext so ausgewählt, dass sie möglichst häufig erfolgreich abgeschlossen und mit mittlerer Güte erkannt wurden. Zusätzlich wurde die Größe und phonetische Ähnlichkeit des Wortschatzes reduziert, um negative Systembewertungen aufgrund von Fehlerkennungen zu minimieren. Die Kommandos beinhalteten das Starten eines Telefonanrufs mit verschiedenen Kontakten und die Navigation zu Einträgen des Telefonbuchs. Eine Übersicht der Kommandos liefert Anhang 12.4.3.

Da die Manipulation des Dialogdesigns nur die abschließende Nachfrage des Systems betraf, hätte eine schrittweise Eingabe den Effekt des flexiblen SGC minimiert. Der Ablauf der Untersu-

chung wurde daher so gewählt, dass die Kommandos den Versuchspersonen aus dem Versuchsleiterraum vollständig vorgegeben wurden und so eine Eingabe mit allen relevanten Informationen provoziert wurde. Eine exakte Wiederholung der Kommandos war zwar nicht explizit gefordert, allerdings konnte durch dieses Vorsprechen ein schrittweises Füllen der Formularfelder vermieden werden.

Im folgenden Absatz werden die beschrieben Dialog- und Feedbackbedingungen in einem Versuchsdesign zusammengeführt.

6.2.2.2 Versuchsplan

Für die Datenerhebung wurde ein gemischtes Versuchsdesign gewählt. Jeder Proband wurde zufällig einer visuellen Feedback-Bedingung zugewiesen. Diese unterschieden sich danach, ob zusätzlich zu dem Zustandsfeedback (graue Symbole) auch die Systemsicherheit farblich (farbige Lautsprecher) vermittelt wurde. Innerhalb der jeweiligen Feedbackbedingung erlebten alle Personen sowohl das statische als auch das dynamische Dialogverhalten in randomisierter Reihenfolge.

Aus der Kombination des zweistufigen Zwischensubjektfaktors Feedback und des ebenfalls zweistufigen Innersubjektfaktors Dialogverhalten, ergibt sich ein 4-Felder-Versuchsplan (siehe Abbildung 40).

Abbildung 40: Vier Versuchsbedingungen der Studie II

6.2.2.3 Erhebungsinstrumente

Zur Erfassung der allgemeinen Gebrauchstauglichkeit fand erneut der standardisierte Usability-Fragebogen **SUS** Anwendung (siehe Kapitel 6.1.2.4). Der Fragebogen wurde um eine subjektive

Einschätzung der **Ablenkung** über das Item *Ich fühlte mich von der Fahraufgabe abgelenkt.* ergänzt.

Wie bereits in der ersten Untersuchung wurde der absolute SUS-Score durch eine Befragung ergänzt, die detailliertere Interpretationen ermöglichen soll. Entsprechend wurden subjektive Maße für Effektivität, Effizienz, Zufriedenheit und weitere Systemcharakteristika konstruiert. Durch diese sogenannten **KU-Ratings** sollten die spezifischen Facetten der Usability und wesentliche Systemeigenschaften detailliert erfassbar gemacht werden. Dazu wurde sich abermals der in Kapitel 6.1.2.4 eingeführten Kategorienunterteilungsskala nach Heller (1985) bedient. Um die unterschiedliche Beurteilung der Probanden aufgrund ihres individuellen Bezugssystems zu minimieren, wurde in Anlehnung an Vollrath und Krems (2011) eine verbale Verankerung der Pole „sehr schlecht“ und „sehr gut“ gegeben (siehe Anhang 12.4.4.2). Jene Verankerung, die beschreibt, was mit den Skalenenden gemeint ist, etabliert ein standardisiertes Bezugssystem über alle Probanden hinweg und soll helfen, die Methodenvarianz zu verringern und die Zuverlässigkeit der Messung zu erhöhen. Darüber hinaus wurden die Probanden extraspektiv instruiert, das heißt, der Versuchsleiter fragte z. B. wie angemessen die Dialoglänge war, nicht wie sie empfunden wurde. Die Befragung erfolgte durch ein hoch-standardisiertes Interview, bei dem den VP die KU-Skala stets vorlag.

Um eine eventuelle Lernkurve abzubilden, wurde jedes System von den Versuchspersonen zweimal hinsichtlich der KU-Skalen bewertet, jeweils nach der Hälfte und am Ende der Kommandos. Der SUS-Fragebogen wurde lediglich nach der Beendigung der Aufgabenfahrten beurteilt.

Im Rahmen des soziodemografischen Teils des Fragebogens fand auch eine Abfrage der **Nutzungsbereitschaft** (engl. *Intention to Use*; ITU) für die Sprachbedienung statt. Dazu wurden sechs Items theoretisch abgeleitet, die die Einstellung der Probanden zu Sprachbedienung und eine verhaltensbezogene Akzeptanz erfragen sollten. Die Nutzungsbereitschaft wird dabei als zentrale Dimension betrachtet, die durch die Aspekte Anstrengung, Fehleranfälligkeit und Modalitätenvergleich erfasst werden soll. Die Items des ITU gestalten sich dabei wie folgt:

- Ich halte Sprachbedienung für eine sinnvolle Alternative zu Tasteneingabe.
- Sprachbedienung strengt mich an.
- Ich kann mir gut vorstellen Sprachbedienung zu nutzen.
- Ich finde Sprachbedienung zu fehleranfällig.
- Ich würde Sprachbedienung als Zusatzausstattung wählen.
- Ich finde Sprachbedienung überflüssig.

Zur Beantwortung der Fragen stand den VP eine 5-stufige Likert-Skala zur Verfügung, deren Skalenbeschriftung sich an die des SUS anlehnte (siehe Abbildung 30). Der Fragebogen wurde vor und nach den Aufgabenfahrten erhoben und kann somit eine mögliche individuelle Veränderung der Nutzungsbereitschaft abbilden.

Den objektiven Maßen lagen neben den Fahr- und Blickdaten auch Beobachtungen durch zwei geschulte Versuchsleiter zugrunde. Eine Übersicht der verwendeten **Protokollbögen** ist Anhang 12.4.4 zu entnehmen.

Einerseits wurde notiert, wie häufig das System die Auslösung falscher Aktionen ohne Bestätigungsanfrage ankündigte. Dies war immer dann der Fall, wenn Name, Funktion (anrufen/navigieren) oder Kategorie (privat/im Büro/mobil) nicht korrekt in die Aktion aufgenommen wurden. Überflüssige Bestätigungsanfragen (engl. *final confirmations*) wurden hingegen gezählt, wenn das System bei Bestätigungsrückfragen alle relevanten Feldinformationen korrekt wiedergab und damit eine entbehrliche Bestätigung erfragte. Auch die Fehlerkennungen, die das System produzierte, wurden kommandoweise vermerkt. In Ergänzung dazu wurden spezifische Fehlbedienungen der Versuchspersonen erfasst. Insbesondere das zu frühe Sprechen eines Kommandos (TTE) wurde durch die VL erfasst. Das Zählen der Turn Takes wurde als Häufigkeitsmaß für die Dialoglänge zusätzlich durch die aufgezeichneten Bedienzeiten abgesichert. Ebenfalls wurde erfasst, ob ein Kommando zu einem erfolgreichen Dialogabschluss führte.

Ferner wurde vermerkt, wenn Nutzer ihre Äußerungen hyperartikuliert einsprachen (vgl. Kapitel 2.2.3). Da die prosodischen Eigenschaften der Hyperartikulation viele Parallelen zu denen der Frustration und des Ärgers (Ang et al. 2002) aufweisen, bietet sich an dieser Stelle die Möglichkeit die (Un-)Zufriedenheit der Nutzer mittels einer objektiven Beobachtung zu erfassen. Zwei VL protokollierten während des Versuchs wie viele Äußerungen hyperartikuliert eingesprochen wurden. Dabei hielten sie sich in unterschiedlichen Räumen auf, um die gegenseitige Beeinflussung zu minimieren.

Zur Bewertung der Visualisierung erfolgte ein strukturiertes Interview, das dem der ersten Studie nahezu identisch ist. Das Interview endete mit offenen Fragen zu fehlenden oder überflüssigen Rückmeldungen und einer Präferenzbeurteilung. Zur Verständlichkeitsüberprüfung der farblichen Darstellung der Systemsicherheit fand weiterhin eine kurze Abfrage der **mentalen Modelle** der Nutzer statt. Im Rahmen einer freien Assoziation sollte überprüft werden, ob die Nutzer durch die Verwendung der Ampelmetapher die Analogie zur Systemsicherheit herstellen können. So wurden den Versuchspersonen nach den beiden Aufgabenfahrten die entsprechenden Icons gemäß ihrer Feedbackbedingung (grau oder farbig) vorgelegt und offen danach gefragt, wofür die Visualisierungen im Dialog standen. Zusätzlich wurde den Probanden der Zustandsfeedbackgruppe die

farbigen Anzeigen gezeigt und ihre Assoziation erfragt. Somit sollte die intuitive Verständlichkeit der Icons geprüft werden.

Eine Übersicht aller Erhebungsinstrumente liefert der Versuchsablauf in Anhang 12.4.

6.2.2.4 Stichprobe

Am Versuch nahmen ausschließlich Mitarbeiter der Volkswagen AG teil. Für die Versuchsteilnahme wurde der Besitz eines Führerscheins, Deutsch als Muttersprache sowie ungestörtes Farbsehvermögen vorausgesetzt. Entsprechend dieser Kriterien wurden 25 Personen aus der Datenbank des betriebsinternen Probandenpools rekrutiert. Zwei Personen wurden aufgrund von Nichterscheinen, bzw. Simulatorübelkeit ausgeschlossen, sodass insgesamt eine Stichprobe von N= 23 Personen zur Verfügung stand.

Die Altersspanne betrug 21 bis 56 Jahre (*MW*= 39.5 Jahre; *SD*= 9.5 Jahre). Mit 11 weiblichen und 12 männlichen Probanden wurde eine annähernde Gleichverteilung der Geschlechter erreicht. Tabelle 28 stellt die Verteilung der Versuchspersonen-Parameter auf die Zwischensubjektbedingungen dar.

Tabelle 28: Stichprobenaufteilung nach Feedbackbedingung

		Alter				Geschlecht	
	N	***Minimum***	***Maximum***	***Mittelwert***	***SD***	***männlich***	***weiblich***
grau	11	29	50	38.73	7.17	6	5
farbig	12	21	56	40.17	11.43	6	6

6.2.2.5 Durchführung

Vor Studienbeginn wurden die Versuchsteilnehmer informiert, dass es sich bei der geplanten Untersuchung um einen Simulatorversuch zur Bewertung eines neuartigen Sprachdialogsystems zur Telefon- und Navigationsbedienung handele und über die Aufzeichnung von Fahr- sowie Blickdaten in Kenntnis gesetzt. Es folgte die soziodemografisch Befragung, in der auch Erfahrungswerte und die Basisnutzungsbereitschaft für SDS erfragt wurden. Im Anschluss daran wurden alle Teilnehmer standardisiert für Fahr- und Bedienaufgabe instruiert und das Blickerfassungssystem kalibriert.

Innerhalb der Instruktion in das Sprachdialogsystem erfolgte keine explizite Erklärung des unterschiedlichen Dialog- und Feedbackverhaltens. Allen Probanden wurde das Beispielkommando *Ich möchte Gerhard Schulze mobil anrufen.* vorgeführt, dass einen ersten Einblick in das einge-

stellte Systemkonzept geben sollte. Sie wurden explizit darauf hingewiesen, dass sie zwei Systeme erleben werden. Vor Beginn der zweiten Erhebungsphase wurde das neue System anhand des gleichen Beispielkommandos eingeführt. Die standardisierte Instruktion kann dem Anhang 12.4.2 entnommen werden.

Jeder Teilnehmer absolvierte eine Eingewöhnungsfahrt, zwei Aufgabenfahrten, sowie Referenzfahrten vor und nach den Aufgabenfahrten. Während der Aufgabenfahrten wurden zehn Sprachkommandos (siehe Kap. 6.1.2.2.2) zum Vergleich der vier Dialogbedingungen ausgeführt. Die Erhebung dauerte durchschnittlich 60 Minuten. Ihr schematischer Ablauf kann Anhang 12.4.1 entnommen werden.

6.2.3 Ergebnisse

Gemäß den Forschungshypothesen wird durch die Manipulation des Dialogverhaltens ein positiver Effekt auf die Gebrauchstauglichkeit erwartet. Daher liegt der Fokus der Analysen auf dem Dialogverhalten. Begleitend wird bei jeder Analyse die unterstützende Funktion des Zwischensubjektfaktors Farbe (Interaktionseffekt) geprüft. Als zweiter Zwischensubjektfaktor wurde die Präsentationsreihenfolge kontrolliert, um trotz randomisierter Zuweisung mögliche Reihenfolgeeffekte zu beachten.

Zur Auswertung des balancierten, gemischt-faktoriellen Designs (Kap. 6.2.2.2) fand das Split-Plot-Vorgehen Anwendung. Praktisch lässt sich dies durch eine ANOVA mit Messwiederholung innerhalb der Zwischensubjektfaktoren darstellen (Jones & Nachtsheim, 2009). Alle Analysen erfolgten erneut mit einem zugrunde gelegten Signifikanzniveau von $\alpha = .05$. Ein Trend wird immer dann berichtet, wenn $p < .10$. Zur Prüfung der Voraussetzungen sei auf Kapitel 6.1.3. verwiesen.

Durch die Verwendung der Kategorienunterteilungsskala nach Heller (Kap. 6.1.2.4) und die kontinuierlich gelabelte Likert-Skala der SUS und des ITU kann ein Intervallskalenniveau für die subjektiven Ratings angenommen werden. Auch für die protokollierten Häufigkeiten der Nutzer-System-Interaktion kann dieses Skalenniveau vorausgesetzt werden.

Bevor die Effekte des Dialogverhaltens und der Feedbackdarstellung auf die Usability-Aspekte berichtet werden sollen, soll durch eine Betrachtung der Anzahl überflüssiger Bestätigungsaufforderungen die Manipulation der Interpretationsstrategie und damit des Dialogverhaltens überprüft werden.

Dazu wurde eine ANOVA mit Messwiederholung durchgeführt. Der Innersubjektfaktor Dialogverhalten zeigte hierbei einen signifikanten Haupteffekt ($F[1,22] = 76.7$; $p < .0001$, $\eta^2 = .78$). So erlebte

jeder Proband innerhalb der statischen Bedingung durchschnittlich 1.14 überflüssige Bestätigungsanfragen pro Kommando, während es in der dynamischen Bedingung nur 0.02 überflüssige Bestätigungsanfragen pro Kommando waren.

6.2.3.1 Effektivität, Effizienz und Zufriedenheit

Zur Bestimmung der **Effizienz** wurde die durchschnittliche Dialoglänge und die Anzahl der Turn Takes als objektive Maße gewählt, die den Gesamtaufwand des Kostenfaktors Zeit bis zur Zielerreichung abbilden. Dabei wurden die durchschnittliche Anzahl der Sprecherwechsel und Dauer über alle Kommandos pro Dialogbedingung summiert. Ebenso wurden die Einschätzungen der Dialoglänge innerhalb der KU-Ratings als subjektives Maß erfasst, die die Verhältnismäßigkeit von Effektivität und nutzerseitigem Aufwand erfassen soll.

Das Dialogverhalten zeigte auf das Konstrukt der Effizienz einen signifikanten Einfluss (F[3,17]= 10.6; p< .0001; η^2= .65). Dieser Effekt zeigte sich sowohl bei der durchschnittlichen Länge der Dialoge (F[1,19]= 27.0; p< .0001; η^2= .59) als auch bei der Anzahl der Turn Takes (F[1,19]= 30.7; p< .0001; η^2= .62). So war die durchschnittliche Länge bis zur Kommandoausführung in der statischen Bedingung fast doppelt so lang wie in der dynamischen (siehe Abbildung 41).

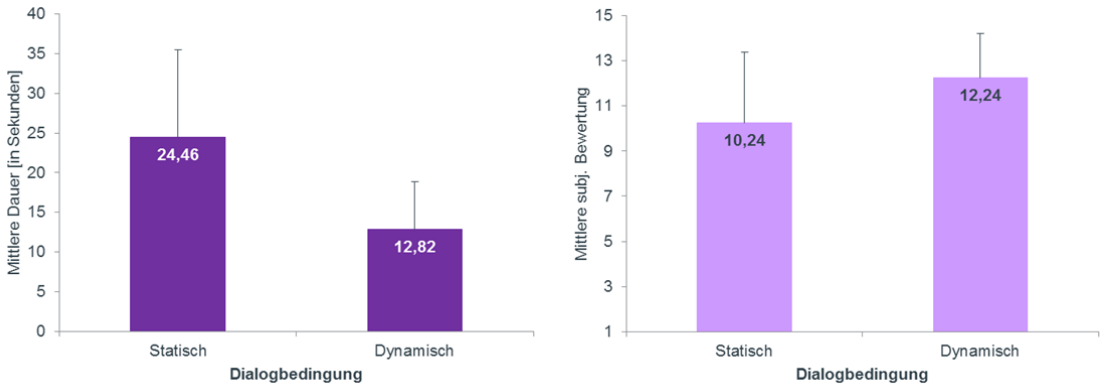

Abbildung 41: Durchschnittliche Dialogdauer (links) und empfundene Angemessenheit der Dialoglänge (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Dialogbedingung

In der Bedingung mit dem liberalen SGC benötigten die VPs durchschnittlich 12.8 Sekunden und 1.5 Turn Takes bis zur Ausführung eines Kommandos, während es in der konservativen Bedingung 24.5 Sekunden und 2.9 Turn Takes waren.

Die Verringerung der Dauer und Sprecherwechsel spiegelte sich auch in der subjektiven Bewertung der Versuchspersonen wider (F[1,19]= 6.0; p= .03; η^2= .24). Während die empfundene Angemessenheit der Dialoglänge in der statischen Bedingung *mittel* bis *gut* eingeschätzt wurde, so

wurde sie in der dynamischen Bedingung an der Grenze zum *sehr gut*en Bereich verortet (siehe Abbildung 41). Es zeigten sich weder Farb- noch Reihenfolgeeffekte auf das Konstrukt Effizienz (Farbe: F[3,17]= 1.2; p= .35 Reihenfolge: F[3,17]= 1.5; p= .25).

Im Rahmen der **Effektivitätsbeurteilung** lässt sich die Vollständigkeit der Zielerreichung vor allem durch die Summe der erfolgreich beendeten Dialoge abbilden, während bei der Genauigkeit zwei Arten von Fehlern Beachtung finden. So wurden zur Erfassung der Effektivität neben den erfolgreichen Dialogabschlüssen, auch die Fehlbedienungen (zu frühe Spracheingaben) und Systemfehler (Fehlerkennung, falsche Aktion) herangezogen. Dabei wurden die Protokollmaße über alle Kommandos pro Dialogbedingung aufsummiert. Die Summe verfrühter Eingaben (TTE) und die Fehlerkennungen (FE) wurden an der Summe der Turn Takes relativiert, da mehr Sprachbeiträge auch beidseitig mehr Fehlermöglichkeiten boten und sie somit direkt von der Manipulation beeinflusst würden. Zusätzlich wurde die subjektive Effektivität über das KU-Rating der wahrgenommenen Zuverlässigkeit der Kommandoausführung erfragt.

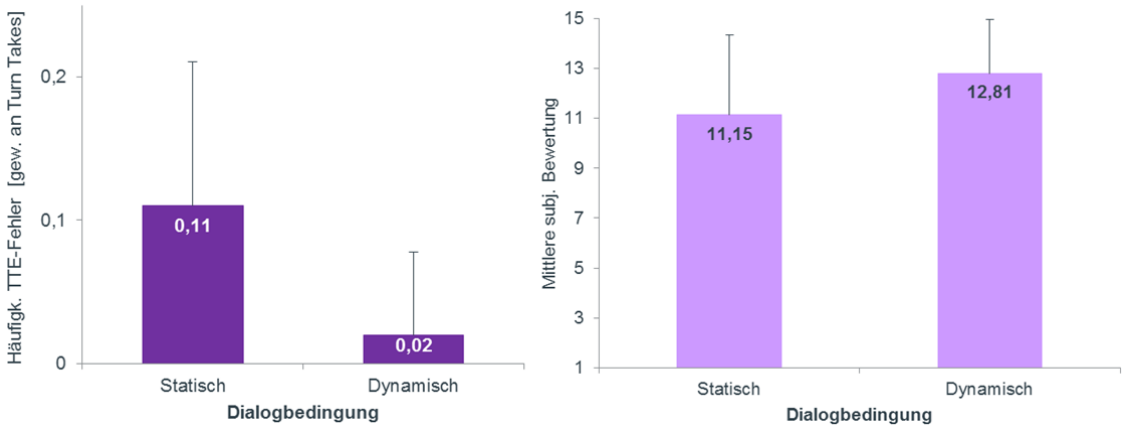

Abbildung 42: Gewichtete Häufigkeit der Fehlbedienungen (links) und wahrgenommene Zuverlässigkeit (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Dialogbedingung

Für das Effektivitätskriterium zeigte sich ein signifikanter Haupteffekt des Innersubjektfaktors Dialogverhalten (F[5,17]= 3.5; p= .03; η^2= .50). Dieser ging maßgeblich auf Gruppenunterschiede in dem Bereich Systemfehler (falsche Aktion), Fehlbedienung (TTE) und subjektiv bewertete Zuverlässigkeit des Systems zurück.

Die subjektiven Zuverlässigkeitsbeurteilungen des Systems wurden allgemein als *gut* eingestuft (siehe Abbildung 42). Dabei zeigte sich dennoch eine positivere Bewertung des dynamischen Verhaltens (F[1,21]= 7.5; p= .01; η^2= .26). Auch unter Berücksichtigung der häufigeren Sprecherwechsel kam es bei festen Rückfragen vor der Kommandoausführung häufiger zu TTE-Fehlern als bei dynamischem Dialogverhalten (F[1,21]= 13.9; p< .0001; η^2= .40) (siehe Abbildung

42). Bei den Fehlerkennungen (gewichtet an der Anzahl der Turn Takes) und in der Vollständigkeit der Zielerreichung konnten keine signifikanten Unterschiede festgestellt werden (FE: F[1,21]< 1; p= .72; Erfolg: (F[1,21]< 1; p= .37). Deskriptive Betrachtungen zeigten, dass sich die Anzahl der erfolgreichen Dialogabschlüsse in beiden Dialogbedingungen 10 annäherte, was einer Abschlussquote von 100% entsprach.

Betrachtet man die falschen Aktionen, die durch das System ausgelöst wurden, so ergab sich eine signifikant größere Anzahl Systemfehler bei dem dynamischen Dialogverhalten (F[1,21]= 6.3; p= .02; η^2= .23). Es zeigten sich keine signifikanten Effekte der farbigen Vermittlung der Systemsicherheit (F[5,17]< 1; p= .67).

Die **Zufriedenheit** unterschied sich von den zuvor untersuchten Parametern Effektivität und Effizienz dahingehend, dass sie sich vorrangig durch subjektive Skalierungsverfahren und Verhaltensbeobachtungen in quantitativen Kennwerten abbilden ließ. Entsprechend wurden hier die subjektive Zufriedenheitsbeurteilung auf der KU-Skala und die protokollierte Verhaltensweise Hyperartikulation den Analysen zugrunde gelegt. Es zeigte sich ein Haupteffekt des Dialogverhaltens auf das Konstrukt Zufriedenheit (F[2,20]= 3.9; p= .04; η^2= .28).

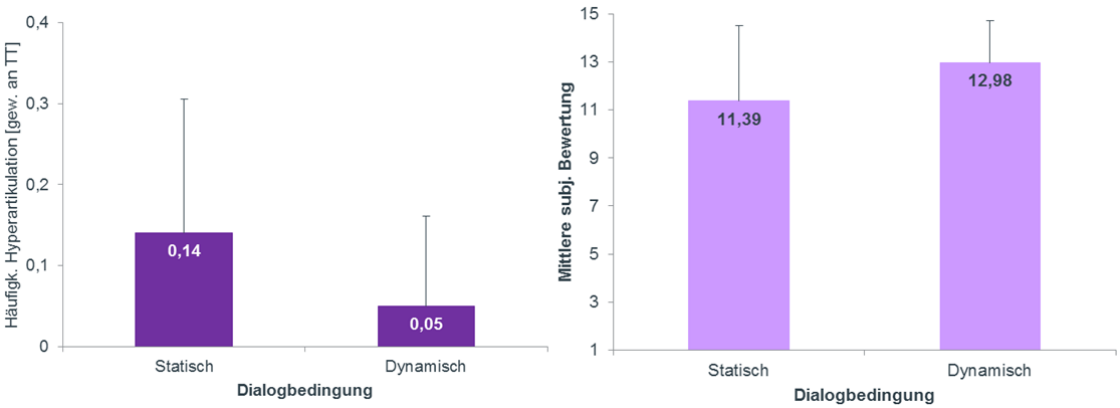

Abbildung 43: Gewichtete Häufigkeit hyperartikulierter Eingaben (links) und Zufriedenheitsbeurteilung (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Dialogbedingung

Dieser Effekt leitete sich aus einem Einfluss der Dialogbedingung auf die subjektive Zufriedenheitsbeurteilung (F[1,21]= 5.4; p= .03; η^2= .21) und auf die Hyperartikulation (F[1,21]= 6.1; p= .02; η^2= .22) her (siehe Abbildung 43). Somit waren Versuchspersonen signifikant zufriedener mit dem dynamischen System und zeigten weniger hyperartikulierte Äußerungen. Erneut zeigten sich keine Effekte der Feedbackbedingung (F[1,21]= 1.7; p= .21) oder der Reihenfolge (F[1,21]< 1; p= .91).

Mit Hilfe einer multiplen linearen Regression wurde ein Prädiktionsmodell für die Zufriedenheitsbeurteilung gerechnet (siehe Tabelle 29). Dabei wurden alle übrigen KU-Ratings der Systembeurteilung als mögliche Prädiktoren schrittweise einbezogen. Die Zuverlässigkeit und die Angemessenheit der Dialoglänge konnte die Zufriedenheit mit dem System vorhersagen (korrigiertes R^2=.74).

Tabelle 29: Prädiktionsmodell für Zufriedenheit

	Effekt Zufriedenheit		
	β	***T***	***p***
Zuverlässigkeit	**.63**	**6.5**	**< .0001**
Dialoglänge	**.33**	**3.4**	**< .01**

Zur Betrachtung der allgemeinen **Gebrauchstauglichkeit** fand erneut der SUS-Score Anwendung. Der Gesamtscore der System Usability Scale lag in allen vier Versuchsbedingungen über 80 Punkten und somit oberhalb des 90. Perzentils (siehe Abbildung 44).

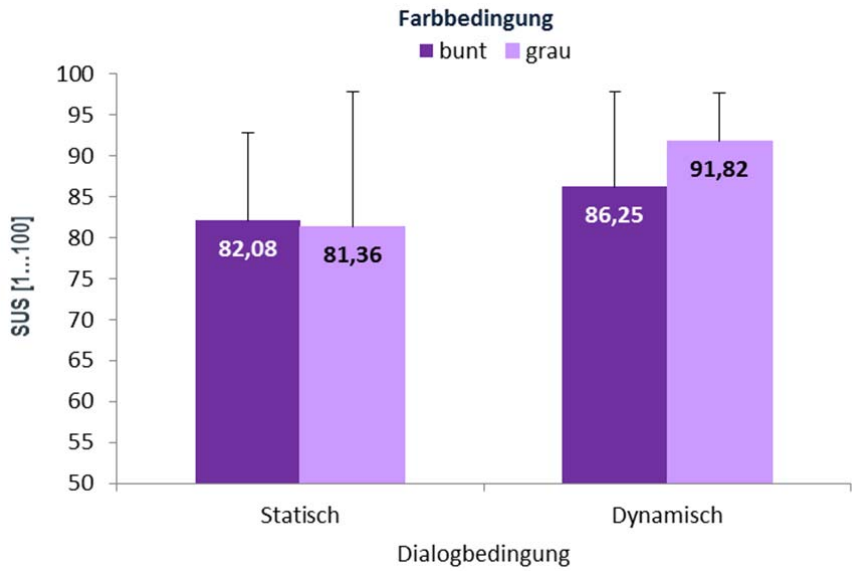

Abbildung 44: SUS-Score in Abhängigkeit der vier Dialogbedingungen

Trotz des hohen Niveaus in allen Bedingungen zeigte sich auch hier ein signifikanter Haupteffekt des Dialogverhaltens. In der dynamischen Dialogbedingung wurden signifikant höhere Gesamtwerte erlangt (F[1,20]= 6.0; p= .02; η^2= .23). Die Zwischensubjektfaktoren zeigten keinen signifikanten Einfluss (Feedbackbedingung: F[1,20]< 1; p= .55; Reihenfolge: F[1,20]< 1; p= .76). Eben-

so konnte keine Wechselwirkung zwischen dem Dialogverhalten und der Feedbackbedingung beobachtet werden (F[1,20]= 1.1; p= .32).

Regressionsanalysen konnten belegen, dass der Effekt der Turn Takes auf den SUS-Gesamtwert durch die Anzahl der Fehlerkennungen nicht vollständig vermittelt werden kann. Im Gegenteil, allein der Effekt der TT wird bei einer integrierten Betrachtung signifikant.

Bei der abschließenden Frage, für welches der beiden Systeme sie sich entscheiden würden, wählten 73.9% der Versuchspersonen das System mit dem dynamischen Dialogverhalten. Nur 6 der 23 VP entschieden sich für das statische Dialogverhalten. Dabei zeigte sich ein Effekt der Präsentationsreihenfolge. Das dynamische Dialogverhalten wurde weniger häufig präferiert, wenn die VP zuerst mit dem statischen System konfrontiert waren (χ^2 [1, N= 23]= 4.1, p= .04). Das Feedbackverhalten zeigte dabei keinen sign. Einfluss auf die Präferenz (χ^2 [1, N= 23]= 0.02, p= .90).

Bei Betrachtung der Anzahl hyperartikulierter Äußerungen zeigten sich signifikante Korrelationen mit der Zufriedenheits- und Zuverlässigkeitsbeurteilung (r= .31, p= .34; r= .38, p= .01). Weiterhin konnte ein signifikanter Zusammenhang mit der Anzahl überflüssiger Bestätigungsanfragen (r= .40, p= .01) und ein Trend mit dem SUS-Score (r= -.28, p= .06) beobachtet werden. Diese Zusammenhänge entfallen allerdings, wenn man das an der Anzahl der Sprecherwechsel gewichtete Maß der hyperartikulierten Äußerungen verwendet.

6.2.3.2 Nutzungsbereitschaft

Die Dialogbedingung, die in diesem Fall auch die Prä-Interaktions-Messung beinhaltete, zeigte auf das Konstrukt der Nutzungsbereitschaft einen signifikanten Einfluss (F[12,68]= 2.8; p< .0001; η^2= .33). Dieser ging auf Gruppenunterschiede in dem Bereich Anstrengung (*Sprachbedienung strengt mich an.*) und Fehleranfälligkeit des Systems (*Sprachbedienung ist zu fehleranfällig*) zurück[26]. Bei beiden Items zeigte sich, dass nach den Systeminteraktionen positivere Bewertungen vorgenommen wurden (Anstrengung: F[1.3,40]= 9.6; p< .01; η^2= .32; Fehleranfälligkeit: F[1.4,40]= 8.7; p< .0001; η^2= .30). Dabei konnte dieser Anstieg der Nutzungsbereitschaft unabhängig von dem Systemverhalten (statisch oder dynamisch) beobachtet werden (F[12,9]= 1.57; p= .25). Es zeigten sich weder auf das Konstrukt Nutzungsbereitschaft noch auf die einzelnen

[26] Da bei nahezu jedem Item die Sphärizitätsannahme verletzt wurde, fand die Greenhouse-Geisser-Korrektur Anwendung.

Variablen Farb- oder Reihenfolgeeffekte (Feedback: F[6,15]< 1; p= .53; Reihenfolge: F[6,15]< 1; p= .75).

6.2.3.3 Gewöhnungseffekt

Wie bereits erläutert, bewerteten die Probanden jedes System zwei Mal in Bezug auf die KU-Items, jeweils nach fünf und nach zehn Kommandos. Für die zwei Erhebungszeitpunkte innerhalb der Dialogbedingungen zeigte keines der KU-Items eine signifikante Differenz (siehe Tabelle 30). Daher beziehen sich die im Folgenden berichteten Effekte auf die über beide Erhebungszeitpunkte gemittelten Bewertungen.

Tabelle 30: Gewöhunungseffekte der Systembewertung

	Statisches System		**Dynamisches System**	
	$F_{(1,22)}$	p	$F_{(1,22)}$	p
Konsistenz	0.6	.45	> 0.1	.90
Transparenz	0.3	.58	0.3	.58
Intuitivität	1.1	.31	> 0.1	.81
Verständlichkeit	2.8	.11	0.3	.58

6.2.3.4 Transparenz- und Konsistenzbeurteilung

Zur Beurteilung spezifischer Systemeigenschaften wurden die KU-Einzelitems zur Erfassung der Transparenz und der Konsistenz des Systems herangezogen.

Es zeigte sich ein signifikanter Interaktionseffekt des Dialogverhaltens und der Feedbackbedingung auf die Transparenzbeurteilung (F[1,19]= 7.0; p= .02; η^2= .27) (siehe Abbildung 45). Während die farbige Vermittlung der Systemsicherheit in der dynamischen Dialogbedingung zu einer signifikant besseren Einschätzung führte als das einfarbig graue Zustandsfeedback, verhielt es sich in der statischen Bedingung genau umgekehrt.

Es konnte darüber hinaus ein signifikanter Haupteffekt für den Zwischensubjektfaktor Reihenfolge festgestellt werden (F[1,19]= 6.7; p= .02; η^2= .26). So wurde die Transparenz in beiden Dialogbedingungen signifikant höher bewertet, wenn die Versuchspersonen zuerst mit dem dynamischen System konfrontiert wurden.

Bei der Konsistenz zeigte sich lediglich ein Trend für den Zwischensubjektfaktor Reihenfolge (F[1,19]= 4.0; p= .06; η^2= .18). Auch hier wurde die Konsistenz in beiden Dialogbedingungen höher bewertet, wenn die VP zuerst das dynamische Dialogverhalten erlebten.

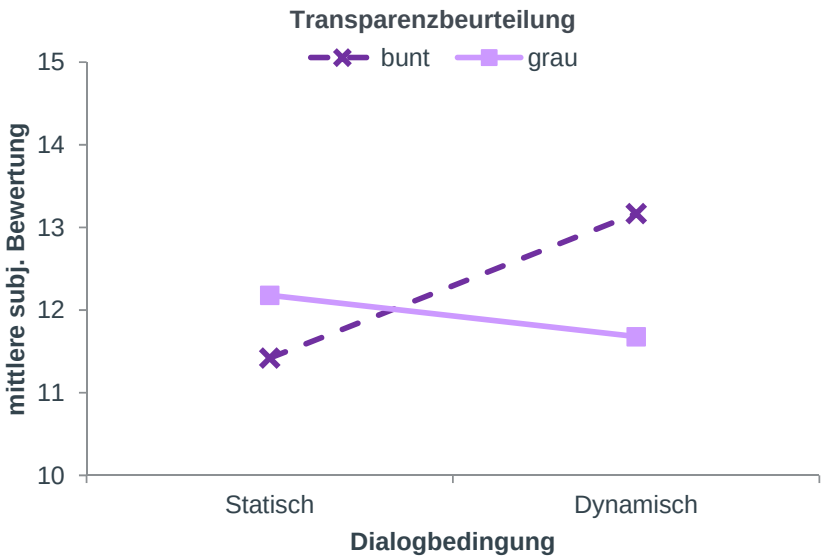

Abbildung 45: Transparenzbeurteilung (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Dialogbedingung

6.2.3.5 Mentale Modelle der Visualisierung

Es ließ sich kein Effekt der Dialog- oder Feedbackbedingung auf das Konstrukt **Gestaltung** beobachten, welches sich aus einer Bewertung der Verständlichkeit, der Größe/Lesbarkeit, der Attraktivität und einer allgemeinen Anzeigenbewertung zusammensetzte (Dialogbedingung: F[4,16]= 4.7; p= .76; Feedbackbedingung: F[4,16]< 1; p= .62). Unabhängig von den Inner- und Zwischensubjektfaktoren bewerteten die Versuchspersonen die Attraktivität, die Größe/Lesbarkeit und die Verständlichkeit der Anzeigen mit *gut*.

Im Rahmen der Auswertung der mentalen Modelle wurde sich auf eine kategorienbasierte Auswertung der Antworten der Versuchspersonen bezogen. Dabei wurden die Antworten dichotomisiert und in korrekt bzw. falsch eingeteilt. 11 Versuchspersonen erlebten die einfarbige Feedbackbedingung, in der ausschließlich das Zustandsfeedback in grau vermittelt wurde. Konfrontiert mit den offenen Fragen, wofür die Mikrofonvisualisierung stand, konnten 100% der VP die richtige Antwort geben. Diese Häufigkeiten können unabhängig davon betrachtet werden, ob sie mit dem statischen oder dynamischen System interagierten. 10 von 11 Versuchspersonen ordneten

auch dem (grauen) Lautsprecher in der statischen Bedingung die richtige Funktion zu, während es in der dynamischen Bedingung sogar 100% waren.

Versuchspersonen, die keine farbige Systemsicherheitsanzeige erlebten, boten eine optimale Stichprobe um die intuitive Verständlichkeit der Ampelmetapher zu überprüfen, da sie unvoreingenommen gegenüber Bewertung der Systemsicherheitsanzeigen sind. Gemittelt über beide Dialogbedingungen ordneten 8 von 11 VP dem grünen und dem orangen Lautsprecher intuitiv die richtige Funktion zu, während nur sieben VP die Funktion des gelben Lautsprechers korrekt antizipieren konnten. Versuchspersonen, die das dynamische Dialogverhalten an zweiter Position erlebten, waren bei dieser Aufgabe schlechter. Ein angeschlossener Test zeigt für die Funktionszuordnung bei dem grünen und orangenen Lautsprecher einen Trend für die Reihenfolge (χ^2= 5.0, *df*=1, *p*= .06)[27].

Die 12 Versuchspersonen, die mit der farbigen Systemsicherheitsanzeige interagierten, zeigten teilweise mehr richtige Antworten. Gemittelt über die Dialogbedingungen wurden der grüne und gelbe Lautsprecher besser beschrieben, wenn die Versuchspersonen die farbige Feedbackbedingung erlebt hatten. Es zeigte sich jedoch kein signifikanter Effekt der Feedbackbedingung auf den Gesamtwert richtiger Antworten (*T*[21]= -0.47; *p*=.65).

Tabelle 31 stellt die prozentuale Anzahl richtiger Antworten in Abhängigkeit von den Dialogbedingungen dar.

Tabelle 31: Prozentuale Anzahl richtiger Funktionszuweisungen

Dialog-bedin-gung	Feed-back-bedin-gung	N	Mikro-fon	Laut-sprecher grau	Laut-sprecher grün	Laut-sprecher gelb	Laut-sprecher orange
statisch	grau	11	100	90.9	72.7	63.6	72.7
dyna-misch	grau	11	100	100			
dyna-misch	farbig	12	91.7	/	100	66.7	58.3
statisch	farbig	12	100	/	83.3	83.3	66.7

[27] Da alle Zellen eine erwartete Häufigkeit kleiner 5 haben, wurde Fishers exakter Test zur Interpretation herangezogen.

6.2.3.6 Fahr- und Blickdaten

Da die Aufzeichnung der Fahrdaten aufgrund technischer Probleme nicht bei jeder Versuchsperson erfolgreich war, belief sich die Stichprobe hier auf 22 Probanden. Die Feedbackgruppen unterschieden sich nicht hinsichtlich ihrer Fähigkeiten zur Spurhaltung während der Referenzfahrt (F[2,19]< 1; p= .39). Zur Analyse der Ablenkung von der Fahraufgabe wurde der Parameter der Spurablage zu vier Messzeitpunkten (Baseline 1, statisches System, dynamisches System, Baseline 2) betrachtet. Im Rahmen der univariaten Varianzanalyse mit Messwiederholung zeigten sich weder Effekte des Dialogverhaltes (F[1.5, 27.8]= 2.1; p= .16), noch der Feedbackbedingung (F[1,19]= 3.0; p= .10) auf die Spurabweichung. Verglich man jedoch nur die beiden Aufgabenfahrten hinsichtlich der Spurhaltung zeigte sich, dass sich die Feedbackgruppen signifikant unterschieden (F[1,19]= 4.9; p= .04; η^2= .21). So war die Abweichung von der Ideallinie bei der alleinigen Darbietung von Zustandsfeedback signifikant größer, als bei der Darbietung des farbigen Sicherheitsfeedbacks.

Die subjektive Einschätzung der Ablenkung zeigte keine signifikanten Effekte des Dialogverhaltens, der Feedbackbedingung oder der Reihenfolge (Dialogbedingung: F[1,19]< 1; p= .44; Feedbackbedingung: F[1,19]< 1; p= .60; Reihenfolge: F[4,16]< 1; p= .93).

Zur Auswertung der **Blickabwendung** fand eine ANOVA mit Messwiederholung innerhalb des Zwischensubjektfaktors Feedback statt. Dabei wurden vier Messzeitpunkte (Baseline 1, statisches System, dynamisches System, Baseline 2) definiert, um einen Anstieg der Blickabwendung während der Aufgabenfahrten zu testen und ggf. Gewöhnungseffekt zu identifizieren. Da die Aufzeichnung der Blickdaten aufgrund technischer Probleme nicht bei jeder Versuchsperson erfolgreich war, belief sich die Stichprobe hier auf 21 Probanden.

Die Blickabwendung war erneut definiert durch die prozentuale Anzahl der Blicke auf das Kombidisplay während der Fahrtzeit (siehe Studie I, Kap. 6.1.3.5). Es zeigte sich ein signifikanter Effekt der Blickabwendung über die Messzeitpunkte (F[2,37.3]= 11.9; p<.0001; η^2= .39). Ausgehend von beiden Baseline-Fahrten ließ sich ein Anstieg bei den Kommandofahrten für beide Systeme beobachten (siehe Abbildung 46). Dabei unterschied sich auch die Baseline 2 signifikant von der Baseline 1 (F[1,19]= 7.5; p= .01; η^2= .28). Im Rahmen der letzten Fahrt ließ sich die geringste Blickabwendung registrieren. Betrachtet man die Innersubjektkontraste, so unterschieden sich die beiden Dialogbedingungen zwar signifikant von den Baselinefahrten (statisch F[1,19]= 8.6; p= .01; η^2= .31; dynamisch F[1,19]= 18.5; p< .0001; η^2= .49), zeigten aber keine signifikanten Differenzen untereinander (F[1,19]= 8.6; p= .01). Innerhalb der Innersubjektkontraste konnte auch ein signifikanter Interaktionseffekt der Feedbackbedingung mit den Messzeitpunkten festgestellt werden (F[1,19]= 3.7; p= .07; η^2= .17). So stieg die Blickabwendung von der Baseline 1 zum statischen System stärker an, wenn dies alleiniges Zustandsfeedback darbot. Bei einer farbigen

Feedbackanzeige fiel der Anstieg geringer aus. Dennoch konnte kein Interaktionseffekt der Blickabwendung mit dem Zwischensubjektfaktor Feedback beobachtet werden (F[2,37.3]= 2.2; p= .13). Ebenso konnte kein Haupteffekt der Feedbackbedingung festgestellt werden (F[1,19]< 1; p= .45).

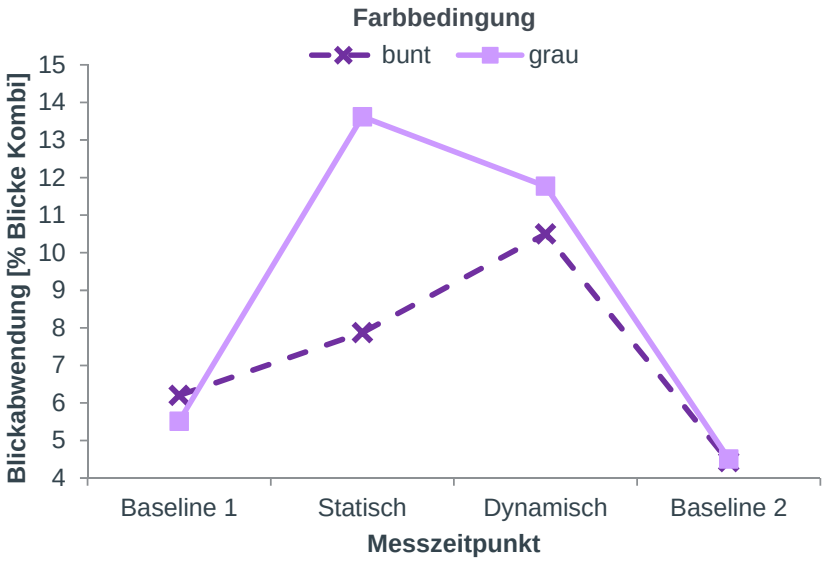

Abbildung 46: Prozentuale Häufigkeit der Blicke auf das Kombidisplay in Abhängigkeit der Dialogbedingungen.

Betrachtete man für die ersten drei Interaktionen die prozentuale Blickdauer je Kommando, so fanden sich teilweise Haupteffekte des Dialogverhaltens (F[3,18]= 4.1; p= .02; η^2= .40). Für das erste und zweite Kommando konnte bei dem dynamischen System eine signifikant höhere Blickabwendung als bei dem statischen System beobachtet werden.

Es zeigte sich kein Interaktionseffekt des Dialog- und Feedbackverhaltens (F[3,18]< 1; p= .55). Für das zweite und dritte Kommando konnte ein Haupteffekt des Zwischensubjektfaktors Feedback beobachtet werden (F[1,20]= 5.6; p= .03; η^2= .22; F[1,20]= 6.4; p= .02; η^2= .24). Gemittelt über die Dialogbedingungen konnte für beide Kommandos eine Reduktion der Blickabwendung bei farbiger Vermittlung der Systemzustände beobachtet werden. Diese Effekte verloren sich nahezu vollständig über die Dauer der Interaktionen.

Nachdem die Ergebnisse der zweiten Erhebung dokumentiert wurden, sollen sie nun bezüglich der ursprünglichen Fragestellungen diskutiert werden.

6.2.4 Diskussion

Eingangs wurde diskutiert, ob der Transfer zwischenmenschlicher Kommunikationsstrategien dazu beitragen kann, die Interaktionsmaximen bei dem Dialog mit SDS zu erfüllen. Angelehnt an die Collaborative Theory (Clark, 1996) wurde ein Dialogverhalten im Sinne des kleinsten gemeinsamen Aufwands (Least Collaborative Effort, Kapitel 3.2.1) umgesetzt, das eine flexible Anpassung des Interpretationsstrategie zeigt. Im Folgenden sollen die Ergebnisse der empirischen Untersuchungen eines SDS mit flexiblen System Grounding Criterions bezüglich der ursprünglichen Frage- und Problemstellungen diskutiert werden.

6.2.4.1 Diskussion der Ergebnisse

Zunächst konnte die **Manipulation des Dialogverhaltens** durch die Flexibilisierung des System Grounding Criterions bestätigt werden. In der dynamischen Bedingung näherte sich die Anzahl überflüssiger Bestätigungsaufforderungen null an. Die Flexibilität des Rückfrageverhaltens belegte die Anzahl von durchschnittlich 1.5 Turn Takes in dynamischer Bedingung. Sie zeigt, dass das System nicht konstant reagierte, sondern bei Unsicherheit auch Bestätigung erbat.

Allein durch die Manipulation des Dialogverhaltens war die Mindestanzahl der erforderlichen Turn Takes per Versuchsdesign bereits vorab unterschiedlich definiert. So wurde in der dynamischen Bedingung bei optimalen Bedingungen nur ein Sprecherwechsel benötigt um ein Kommando erfolgreich zu beenden, während es in der statischen Bedingung durch das konstante Anfrageverhalten mindestens zwei Sprecherwechsel benötigte. Da die objektiven Parameter der **Effizienz** somit direkt von der Manipulation betroffen waren, zeigten sich beachtliche Effektstärken für die Dialogdauer und die Anzahl der Turn Takes.

Diese deutliche Effizienzsteigerung spiegelte sich auch in den subjektiven Bewertungen der Nutzer wieder. Durch eine deutliche Verringerung der Dialogschritte reduzierten sich für die Nutzer die Rezeptions- und Wartekosten. Dies hatte zur Folge, dass sie die Dialoglänge in der dynamischen Bedingung, verglichen zur statischen Bedingung, als angemessener beurteilten.

Auch die **Effektivität** der Interaktionen wurde durch das Dialogverhalten beeinflusst. So zeigte sich bei dynamischer Dialoggestaltung eine geringere Anzahl zu früher Spracheingaben durch die Versuchspersonen. Dieser Effekt blieb auch dann bestehen, wenn die Anzahl der TTE-Fehler an den Turn Takes gewichtet wurde. Somit ließ sich ausschließen, dass dieser Effekt allein durch die Möglichkeit zur Fehlbedienung vermittelt wurde. Es lässt sich eher vermuten, dass Bestätigungsaufforderungen, die keine neuen Informationen beinhalten, sondern die Nutzereingabe lediglich wiederholen, eine überlappende Spracheingabe zusätzlich begünstigen und damit Bedienfehler provozieren. Da Nutzer in der statischen Bedingung die Dialoglänge als verbesserungswürdig

empfanden, kann der TTE-Fehler als Versuch interpretiert werden, die Effizienz des Gesprächs selbstständig voranzutreiben.

Hinsichtlich der systemseitigen Fehlerkennungen konnten keine Effekte der Dialoggestaltung nachgewiesen werden, sofern man sie an der Anzahl der Turn Takes relativierte. Insgesamt ließ sich eine sehr geringe Toleranz gegenüber Fehlererkennungen beobachten. Selbst wenn es nur zu einer Fehlererkennung bei zehn Interaktionen kam, so wirkte sich dies stark negativ auf die Systembeurteilung aus.

Die Manipulation des Dialogdesigns hatte auch keinen Einfluss auf die Abschlussquote. Da die Zielerreichung von beiden Systemen zu annähernd 100% gewährleistet wurde, konnte das dynamische Dialogverhalten nicht zur Steigerung der Vollständigkeit der Zielerreichung beitragen. Dieser Deckeneffekt der Dialogabschlussquote ließ sich damit erklären, dass beiden Dialogsystemen der gleiche verbesserte Spracherkenner zugrunde lag und nur ein begrenztes Funktionsspektrum adressiert wurde. Auch das Nachsprechen der von dem Versuchsleiter vorgegebenen Kommandos kann als Grund für hohe Erfolgsrate aufgeführt werden. So bestand für die Versuchspersonen keine Gefahr, ein Kommando zu verwenden, das nicht in dem Wortschatz des Systems vorkommt (OOV-Fehler).

Dennoch zeigte sich ein Effekt des Dialogverhaltens auf die Ankündigung falscher Aktionen. Bedingt durch die Dialoggestaltung kam es durch das statische Nachfrageverhalten nur in Ausnahmefällen dazu, dass der Name, die Funktion (anrufen/navigieren) oder die Kategorie (privat/im Büro/mobil) nicht korrekt in die Aktionsankündigung aufgenommen wurden. Wurde in der dynamischen Dialogbedingung allerdings etwas Falsches mit hoher Konfidenz verstanden, so wurde die Aktion direkt angekündigt, ohne den Nutzer die Eingabe bestätigen zu lassen. Auch wenn solche false positves in der dynamischen Bedingung signifikant häufiger auftraten, so kann die Anzahl dieser falsch angekündigten Aktionen mit durchschnittlich 0.5 von 10 Kommandos als gering eingestuft werden.

Bedeutsam ist vor allem, dass weder die Erfolgsrate noch die subjektive Bewertung der Systemzuverlässigkeit davon negativ beeinträchtigt wurden. Wenngleich sich eine signifikant größere Anzahl Systemfehler (bezüglich der falschen Ankündigungen) bei dem dynamischen Dialogverhalten ergab, so bewerteten die Versuchspersonen das dynamische System als zuverlässiger. Eine mögliche Erklärung dieser Befundlage kann durch die Hinzunahme des adaptierten Contribution Modell von Brennan und Hulteen (1995) geschehen (siehe Kap.4.1.1). So kann davon ausgegangen werden, dass das umgesetzte System maximal Zustand 5 (Das System hat eine Handlungsabsicht.) erreichen konnte und dies durch die Ankündigung der Handlung (*Helga Otto. Es wird gewählt.*) ausdrückte. Das System meldete also lediglich zurück, dass es eine Aktion nun starten würde. Ruftöne oder eine Routenberechnung, die Zustand 6 entsprochen hätte, erfolgten

nicht. So entstand bei den Versuchspersonen der Eindruck, die Aktion an dieser Stelle noch ohne unangenehme Konsequenzen abbrechen oder korrigieren zu können. In Anlehnung an Dix et al. (2004) kann geschlossen werden, dass Fehler des SDS zugunsten einer höheren Effizienz akzeptiert werden und als wenig problematisch gelten, so lange die Möglichkeit zur Korrektur gegeben ist. Entsprechende Reparaturmechanismen sollten deshalb in zukünftigen Systemen stets implementiert werden.

Als alternative Erklärung soll auch die **artifizielle Laborsituation** diskutiert werden. Konfrontiert mit einem künstlichen Telefonbuch ist die Peinlichkeit eines falschen Anrufes nicht vergleichbar mit der Verwendung realer Kontakte (Kap. 6.1.4). Man kann davon ausgehen, dass diese Rahmenbedingung die Verzeihlichkeit der falschen Aktionen erhöhte. So sollten sich Studien anschließen, in denen die persönlichen Kontakte des Nutzers verwendet werden, um Anrufe zu simulieren.

Die schlechtere Zuverlässigkeitsbeurteilung des statischen Systems kann weiterhin dadurch erklärt werden, dass die ermüdenden Bestätigungsanfragen als Missverständnisse oder Systemunsicherheiten fehlinterpretiert wurden. In Anlehnung an die zwischenmenschliche Kommunikation erwarten Nutzer Rückfragen nur dann, wenn sich der Empfänger unsicher ist. Eine inflexible, defensive Interpretationsstrategie wurde von Nutzern als Heuristik zur Schätzung der Systemkompetenz herangezogen und als Systemschwäche (fehl-)interpretiert.

Auch das Konstrukt der **Zufriedenheit** wurde positiv durch die dynamische Dialoggestaltung beeinflusst. So waren Versuchspersonen signifikant zufriedener mit dem dynamischen System und zeigten weniger hyperartikulierte Äußerungen. Durch die hohe Reaktanz der Nutzer auf die Anforderung jede einzelne Eingabe bestätigen zu müssen, ließen sich hyperartikulierte Äußerungen in der statischen Bedingung als Frustrationsindikatoren vermehrt beobachten.

Es sollte allerdings kritisch hinterfragt werden, ob Hyperartikulation als objektives Maß der Zufriedenheit herangezogen werden kann. Einerseits war die Interrater-Reliabilität derart gering, dass man nicht davon ausgehen kann, dass eine valide Einschätzung über Versuchsleiter hinweg geschehen kann[28]. Andererseits zeigten sich Zusammenhänge der Anzahl hyperartikulierter Äuße-

[28] Zur Ermittlung der Beurteilerübereinstimmung wurde eine Intraklassen-Korrelationen (ICC) gerechnet (Shrout & Fleiss, 1979; McGraw & Wong, 1996; Wirtz & Caspar, 2002). Da für jeden Fall von den Ratern Einzelwerte vorlagen und die Rater nicht zufällig ausgewählt wurden, fand nach Shrout & Fleiss (1979) ein ICC (3,1) mit justierten Schätzern Anwendung. Anhang 12.4.6 sind die system- und kommandospezifischen Intraklassen-Korrelationen zu entnehmen.

rungen mit gängigen Usability-Maßen. Sie sind jedoch als gering einzustufen und verschwinden gänzlich, wenn man die Anzahl hyperartikulierter Äußerungen an der Anzahl der Turn Takes misst. Somit scheint dieser Effekt vollständig über die Anzahl der Sprecherwechsel vermittelt.

Zur Betrachtung der allgemeinen **Gebrauchstauglichkeit** sollten die Ergebnisse des Usability-Fragebogens diskutiert werden. In jeder der Versuchsbedingungen wurde ein Wert von 80 Punkten erreicht. Damit lagt der Gesamtscore nicht nur über dem Durchschnitt, sondern auch oberhalb des 90. Perzentils. Er kann also über alle Varianten hinweg als sehr gut bezeichnet werden und liefert einen weiteren Nachweis, dass die Qualität des Sprachdialogsystems auch ohne Manipulation bereits auf einem zufriedenstellenden Niveau vorlag. Dennoch zeigte sich für das dynamische Dialogverhalten, verglichen mit der statischen Bedingung, eine weitere Verbesserung des Gesamtscores.

Insgesamt zeigten die Versuchspersonen eine deutliche **Präferenz** für das dynamische System. Nur 6 der 23 VP entschieden sich für das statische Dialogverhalten. Dabei sollte in Betracht gezogen werden, dass fünf der sechs Versuchspersonen, die sich gegen das dynamische System entschieden, zuerst mit dem statischen System interagierten. Nachdem sie sich an die Rückversicherung des Systems gewohnt hatten, empfanden sie eine sofortige Handlungsankündigung als kritisch und verunsichernd.

Bezüglich der **Gewöhnungseffekte** muss die Nullhypothese beibehalten werden. Es zeigten sich keine Effekte der Erhebungszeitpunkte auf die Systembeurteilungen. So wurde weder das dynamische Dialogverhalten zunehmend besser, noch das statische Rückfrageverhalten zunehmend schlechter beurteilt. Verschiedene Ursachen lassen sich für diese Befundlage diskutieren. Einerseits kann man mutmaßen, dass auch nach 10 Systeminteraktionen kein Gewöhnungs- oder Experteneffekt beobachtbar sein kann. Zur vollständigen Überprüfung dieser Hypothese sollten Langzeituntersuchungen angeschlossen werden. Selbstverständlich kann auch von Memorierungseffekten ausgegangen werden. So lagen zwischen den beiden Erhebungen durchschnittlich nicht mehr als fünf Minuten. Es ist also vorstellbar, dass die Versuchspersonen sich an ihre vorangegangene Bewertung noch gut erinnerten und ähnlich antworteten, um ein möglichst konsistentes Einstellungsbild zu hinterlassen.

Andererseits zeigte sich von Beginn an eine bessere Bewertung des dynamischen Dialogverhaltens. Auch Novizen scheinen diese Art der flexiblen Dialogführung als natürlicher und angenehmer zu empfinden. Besonders hervorzuheben ist, dass sich keine Unterschiede der **Konsistenzbeurteilung** zeigten. Das flexible Rückfrageverhalten wurde genauso konsistent wie das

statische Interpretationsverhalten beurteilt[29]. Auch ohne Vermittlung des Sicherheitsfeedbacks zeigten sich keine negativen Effekte des dynamischen Dialogverhaltens bei der Bewertung der Konsistenz. Es kann geschlossen werden, dass zugunsten der Implementation zwischenmenschlicher Kommunikationsstrategien auf eine hohe Konsistenz der Systemrückmeldungen verzichtet werden kann.

Eine weitere Annahme bezog sich auf die bessere **Transparenzbewertung** des dynamischen Dialogverhaltens, wenn es in Kombination mit der farbigen Vermittlung der Systemsicherheit auftrat. Innerhalb der dynamischen Dialoge führte die Vermittlung des Sicherheitsfeedbacks zu einer besseren Transparenzeinschätzung als die Vermittlung des Zustandsfeedbacks. Dieses Ergebnis belegt, dass die dynamischen Dialoge davon profitierten, wenn ihre mangelnde Konsistenz anschaulich Erklärung fand. Der Effekt des mangelnden Sicherheitsfeedbacks ist in der dynamischen Bedingung so stark, dass das dynamische System mit grauem Zustandsfeedback sogar dem statischen innerhalb der gleichen Feedbackbedingung unterlegen ist. Alleiniges Zustandsfeedback ist somit nicht ausreichend, um das Grounding bei flexibler Dialoggestaltung adäquat zu unterstützen.

Erwartungskonform scheint hier die Feststellung Clarks (1996) zu gelten, dass Dialoge zwei Spuren der Kommunikation (Kap. 4.2.1) benötigen. Eine Erkenner-Mimik zur Vermittlung von Verständnisevidenzen und zur Erklärung des Dialogverhaltens konnte zu einer besseren Systembewertung beitragen. Entsprechend der Annahmen trat dieser Vorteil aber nur in der Systemkonfiguration auf, in der durch die Visualisierung zusätzliche, dem Grounding zuträgliche Informationen vermittelt wurden. Während innerhalb der dynamischen Bedingung durch das farbige Sicherheitsfeedback die Ursachen für variierendes Rückfrageverhalten erklärt wurden, vermittelte dasselbe Feedback in der statischen Bedingung Zusatzinformationen, die nicht zur Fehlervermeidung oder Aufklärung beitragen konnten. Die visuelle Vermittlung der Systemsicherheit scheint somit überflüssig, wenn sie innerhalb des Dialogdesigns keine Beachtung findet. Die redundanten Informationen der farbig-statischen Bedingung erhöhten lediglich den Verarbeitungsaufwand, ohne einen Mehrwert zur Etablierung des GCs liefern zu können. Darauf reagierten die Nutzer sensibel, denn es wurden in diesen Fällen zwei wesentliche Kommunikationsmaximen verletzt, die der Relevanz und der Quantität (Kap. 3.1.2.1). Vor dem Hintergrund dieser Ergebnis-

[29] Lediglich zwei Versuchspersonen erlebten das dynamische System konstant, da bei ihnen jede Eingabe den definierten Konfidenzschwellenwert überschritt. Auch wenn sie aus der Analyse ausgeschlossen wurden, zeigten sich keine Differenzen in der Konsistenzbewertung zwischen dem statischen und dem dynamischen System.

se scheint es bei der Nutzung verschiedener Feedbackkanäle von zentraler Bedeutung zu sein, ihren jeweiligen Informationsbedarf und deren wechselseitige Ergänzung zu beachten. Wie bereits Studie I zeigte, ergeben sich aus Visualisierungen, die einen geringen informativen Mehrwert liefern, keine positiven Effekte.

Im Einklang zu den **Reihenfolgeeffekten** der Präferenzurteile konnte beobachtet werden, dass trotz der randomisierten Zuweisung die Transparenz und Konsistenz in beiden Dialogbedingungen höher bewertet wurde, wenn die Versuchspersonen zuerst mit dem dynamischen System konfrontiert wurden. Eine Verringerung der Informationen (statisch vor dynamisch) führte in erster Instanz zu einer deutlicheren Wahrnehmung der Inkonsistenz der dynamischen Dialoge. Als Gestaltungsempfehlung lässt sich ableiten, dass bei der Umsetzung eines flexiblen SGC von Beginn an ein dynamisches Dialogverhalten gezeigt werden sollte. Von einer Anpassung des Rückfrageverhaltens im Verlauf des Dialogs soll daher abgeraten werden.

Die Transparenz stellte die einzige Variable dar, die durch den Zwischensubjektfaktor Feedbackverhalten beeinflusst wurde. Es ist diskussionswürdig, dass sich bei der System Usability Scale ausschließlich ein positiver Effekt des dynamischen Dialogverhaltens zeigte und kein Interaktionseffekt mit der Konfidenzvisualisierung. Man hätte erwarten können, dass sich die Einflüsse der zusätzlich gesteigerten Transparenz auch im SUS-Gesamtscore niederschlagen. Allerdings wiesen die Verankerungen der Skalenpole des Transparenzurteils keine Parallelen zu den SUS-Formulierungen auf. Der Aspekt der Transparenz, der durch die Vermittlung der Systemsicherheit verbessert wurde, ist dem SUS damit nicht inhärent. Daher zeigte sich auch keine signifikante Beeinflussung des SUS durch die visuelle Feedbackgestaltung. Bezüglich der **grafischen Anzeigen** konnte die Annahme bestätigt werden, dass auch Versuchspersonen, die nur mit der einfarbigen Feedbackbedingung konfrontiert wurden, ähnliche Assoziationen zu den farbigen Icons zeigten wie Konfidenzvisualisierungsgruppen. Dabei zeigten sich bei der Funktionszuordnung zu dem grünen und orangenen Lautsprecher kaum Unterschiede. Beide Gruppen zeigten in einem hohen Ausmaß die richtigen Assoziationen. Die Funktion des gelben Lautsprechers schien für die Gruppe mit alleinigem Zustandsfeedback am schwersten vorstellbar. So wurde dieser häufiger mit der Alternativinterpretation[30] in Verbindung gebracht. Hatten die Versuchspersonen mit dem

[30] Einerseits kam es zu Verwechslungen bezüglich der Funktionszuschreibung. So brachten drei Versuchspersonen die Farbe mit einer Aktion in Verbindung (Anrufen oder Navigation). Andererseits wurde auch eine Verbindung zu dem Aktivitätszustand des Systems fälschlicherweise hergestellt. So wurde Grün mit dem Anschalten des Systems, Gelb mit dem Standby-Zustand und Rot mit einer Stummschaltung erklärt.

farbigen System interagiert, so stieg die Anzahl der richtigen Funktionszuschreibung des gelben Lautsprechers insbesondere in der statischen Bedingung leicht an. Die geringe Anzahl richtiger Antworten bei den gelb- und orangefarbenen Lautsprechern innerhalb der dynamischen Gruppe lässt sich dabei auf die Anzeigehäufigkeit zurückführen. So wurde der orangefarbenen Lautsprecher in beiden Dialogbedingungen nur bei zwei Versuchspersonen angezeigt. Der gelbe Lautsprecher wurde in der dynamischen Bedingung im Durchschnitt nur 3.6mal innerhalb der 10 Kommandos angezeigt. Viele Versuchspersonen sahen ihn nur ein bis zweimal im Verlauf der Untersuchung. In der statischen Bedingung wurde er dagegen, aufgrund der fest implementierten Nachfrage im Durchschnitt 19.8mal innerhalb der 10 Kommandos angezeigt. Keine der Versuchspersonen sah ihn weniger als 10mal. Somit lässt sich auch das bessere Antwortverhalten in der statischen Bedingung erklären.

Dennoch lässt sich eine Quote richtig antizipierter Antworten der Zustandsvisualisierungsgruppe von durchschnittlich 69.7% und eine Quote richtiger Antworten der Konfidenzvisualisierungsgruppe von 76.4% als erfolgreich bewerten. Das trotz des offenen Fragestils und des Fehlens einer expliziten Instruktion in die Visualisierungen eine derartige Quote richtiger Antworte erzielt werden konnte, verdeutlicht die Intuitivität der Anzeigen. Dem Zustandsindikator Mikrofon-Icon wurde in 100% der Fälle eindeutig die richtige Funktion zugewiesen. Auch das graue Lautsprecher-Icon wurde zu 90.9% korrekt benannt.

Bezüglich der **Nutzungsbereitschaft** zeigte sich zwar eine höhere, allgemeine verhaltensbasierte Akzeptanz nach der Systeminteraktion als vorher; dieser Anstieg ist allerdings unabhängig von dem Dialogverhalten. So wurde die Anstrengung durch Sprachbedienung und die Fehleranfälligkeit der Systeme nach der Untersuchung geringer eingeschätzt. Allein die Konfrontation mit dem neuartigen SDS bewirkt eine Steigerung der Nutzungsbereitschaft, unabhängig von der Versuchsbedingung. Es bestand keine Differenz zwischen der Nutzungsbereitschaft des statischen Systems und des dynamischen System. Damit konnte nicht belegt werden, dass die dynamischen Dialoge durch die Implementation zwischenmenschlicher Interaktionsstrategien zu einer höheren, verhaltensbasierten Akzeptanz führen.

Als Ursache für den ausbleibenden Effekt sollte die Formulierung der Items diskutiert werden. Bisher beziehen sich alle Items auf Sprachbedienung allgemein und mögliche Effekte der Dialoggestaltung liefen so Gefahr, durch Memorierungseffekte überlagert zu werden. Für den Vergleich der beiden Postmessungen ist die Formulierung der Items ggf. nicht sensibel genug gewesen. Für spätere Untersuchungen sollten daher die Formulierungen der Items des ITU bei den Post-Interaktions-Messungen spezifischer erfolgen.

Mit Hilfe der Ergebnisse der Fahr- und Blickdatenanalyse konnte die Annahme bestätigt werden, dass weder das dynamische Dialogverhalten noch die ergänzenden Konfidenzvisualisierungen zu einer erhöhten **Ablenkung** führen.

Betrachtete man allerdings die prozentuale Blickabwendung von der Straße für die ersten fünf Kommandos detaillierter, so ließen sich Effekte der Dialog- und Feedbackgestaltung belegen. Während es bei den ersten beiden Kommandos innerhalb der dynamischen Bedingung zu höheren Blickabwendungen kam, zeigte sich ein positiver Effekt der farbigen Visualisierung. Letztere ließen sich vor allem dadurch erklären, dass die Farben und deren Wechsel in der Peripherie besser wahrnehmbar waren, als der Wechsel eines grauen Mikrofons in einen grauen Lautsprecher. Dies erforderte mehr Blickabwendungen und wirkt sich im direkten Vergleich auch negativ auf die Fähigkeit zu Spurhaltung aus.

Der Haupteffekt des Dialogverhaltens lässt sich vor allem dadurch erklären, dass den Versuchspersonen zu Beginn noch Vertrauen zu dem System fehlte und sie sich bei der Aktionsankündigung rückversichern wollten, dass die korrekte Handlung ausgeführt wurde. Dazu schauten sie in der dynamischen Bedingung häufiger auf das textuelle Inhaltsfeedback als in der statischen Bedingung, bei welcher jeder Aktionsbestandteil akustisch rückversichert wurde.

Des Weiteren gab es einen Unterschied in der akustischen Gestaltung der Prompts. Während in der statischen Bedingung jede Einheit akustisch wiederholt wurde, wurde in der dynamischen Bedingung nur der Name im Rahmen der Aktionsankündigung wiederholt (*Helga Otto. Es wird gewählt.*). Der Nummerntyp wurde durch das dynamische System nicht noch einmal vorgelesen und war somit allein dem textuellen Inhaltsfeedback zu entnehmen. Im Rahmen der Navigation wurde die Aktion nur mit *Navigation wird gestartet.* angekündigt, ohne den Kontakt zu dem gefahren werden sollte zu nennen. Dies führte besonders bei den anfänglichen Kommandos zu einer erhöhten Blickabwendung und stellt eine klare Differenz zu dem statischen System dar, bei dem alle Inhalte akustisch vermittelt wurden. Der Effekt belegt, dass es zu Blickabwendung insbesondere dann kommt, wenn die Aussagen des Systems nicht selbsterklärend oder unvollständig sind.

Im Verlauf der Blickdaten über die Kommandos kann jedoch eine Lernkurve verzeichnet werden. Nach fünf Kommandos, als sich die Versuchspersonen der Zuverlässigkeit des Systems sicher waren und die Funktionsweise bekannt war, ließen sich keine Differenzen in der Blickabwendung mehr feststellen. In gewisser Weise können somit die für die Systembewertung erwarteten Gewöhnungseffekte für den objektiven Parameter der Blickabwendung bestätigt werden.

Es wurde keine subjektive Erleichterung der mentalen Belastung durch das dynamische Dialogverhalten von den Versuchspersonen berichtet. Dies hätte erwartet werden können, da nicht nur

die Kosten des Grounding-Prozesses durch das flexible Rückfrageverhalten reduziert werden (siehe Kap. 3.2.1), sondern auch, da sich der zeitliche Aufwand reduziert. Allerdings nahmen die Versuchspersonen die mentale Beanspruchung beider Systeme ähnlich wahr. Das flexible Rückfrageverhalten konnte die Dialoge mit dem SDS nicht in Richtung automatischer Dialogführung verschieben.

Insgesamt zeigte sich, dass die **Effizienz** eine erfolgsversprechende Stellgröße der Dialoggestaltung ist. So ergaben bereits Untersuchungen von Wechsung et al. (2010) und Brumby et al. (2011), dass die Wahrscheinlichkeit der Nutzung einer Modalität davon abhängt, wie viele Interaktionsschritte durchzuführen sind. Durch die Flexibilisierung des Nachfrageverhaltens konnte die Anzahl der nötigen Interaktionen knapp um die Hälfte reduziert werden, ohne die Zuverlässigkeit des Systems zu beeinträchtigen. Im Folgenden gilt es, die Systemumsetzung zu diskutieren.

6.2.4.2 Diskussion Umsetzung

Die vorliegende Untersuchung konnte aufzeigen, dass die **Konfidenz** zur Adaption des System Grounding Criterions als ausreichend bewertet werden kann. Allein anhand der Systemsicherheit des voran gegangenen Dialogschritts wurde entschieden, ob eine explizite Evidenz, z. B. in Form einer Bestätigungsaufforderung, oder implizite Evidenz durch die direkte Aktionsausführung gerechtfertigt waren. Dies führte zu einer beinah vollständigen Eliminierung überflüssiger Bestätigungsanfragen in der dynamischen Dialogbedingung. Obwohl es dabei auch zu falschen Aktionsausführen kam, schienen diese von den Nutzern als verzeihlich eingestuft zu werden. Dennoch sollte die Anwendung des dynamischen Dialogverhaltens bei Langzeitnutzung und unter Verwendung eines realen Telefonbuchs erneut getestet werden. Die Akzeptanz der Fehlankündigungen sollte somit je nach Kontext und Gesamtzusammenhang beleuchtet werden.

Auch Litman und Pan (2002) verwendeten die Erkennerkonfidenz als Prädiktor für Dialogperformance und definierten einen Schwellenwert, ab dem das System ausführlichere oder negative Evidenzen berichtete. Diese Adaptivität verbesserte die Erfolgsrate und die Zufriedenheit der Nutzer mit dem Sprachdialogsystem. Während sie somit eine **Korrektur** in die defensive Richtung nach vier kritischen Dialogschritten vornahmen und die Rückmeldungen situationsabhängig ausführlicher gestalteten, war in der vorliegenden Umsetzung die Zielsetzung Dialoge mit guter Erkennrate zu beschleunigen. Daher zeigten sich keine ähnlich starken Verbesserungen der Erfolgsrate. Eine konstante Erfolgsrate kann jedoch bei der gegebenen Aufwärtskorrektur in der durchgeführten Studie als erfolgreich betrachtet werden. Für eine Weiterentwicklung des vorliegenden Systems wäre es vorstellbar, eine Abwärtskorrektur der Interpretationsstrategie bei problematischen Dialogen, wie Litman und Pan (2002) sie vornehmen, als Ergänzung zu implementieren. Somit würden sich die Interaktionsstrukturen dann flexibel daran orientieren, wenn die vorangegangenen Dialoge sich problematisch oder fehlerfrei gestalteten. Insbesondere nach

vorangegangenen problematischen Interaktionen kann man sich vorstellen, auch positive Rückmeldungen über das Erreichen eines gewissen Dialogzustandes verstärkt zu bekommen.

Weiterhin gilt es anzumerken, dass das vorliegende System bei **schrittweisen Eingaben** kaum eine Verbesserung der Effizienz zeigte. Gaben Nutzer die Informationseinheiten (Aktion, Name, Nummerntyp) schrittweise ein, so konnten die Dialogschritte nicht reduziert werden. Das vorliegende dynamische System entschied anhand der Konfidenz der letzten Eingabe, ob es sich alle Eingaben noch einmal zusammengefasst bestätigen ließ. Bis zu dieser letzten Eingabe waren nur implizite Bestätigungen implementiert. Im schlimmsten Fall hätte dies bedeuten können, dass selbst wenn die ersten beiden Informationseinheiten die definierte Konfidenzschwelle unterschritten hätten und somit keine ausreichende Verständnisgüte vorlag, das System sofort die Aktion angekündigt hätte, sobald die dritte und letzte Einheit mit hoher Konfidenz erkannt wurde. Ein Beispieldialog ist im Folgenden dargestellt:

Nutzer	Anrufen.	
System	Wen möchten sie anrufen?	
Nutzer	Hella von Motto.	Konfidenz< 40%
System	Den Nummerntyp bitte.	
Nutzer	Mobil	Konfidenz> 40%
System	Helga Otto. Es wird gewählt.	

Als Konsequenz würde bei diesem Beispiel eine falsche Feldinformation, trotz geringer Konfidenz unbestätigt übernommen und der falsche Teilnehmer angerufen.

Auch wenn diese schrittweise Eingabe durch das Vorsprechen kompletter Kommandos im Rahmen der Untersuchung vermieden werden konnte, so sollte bei zukünftigen Implementationen der Fall der schrittweisen Eingabe Beachtung finden. So könnten sich spätere Umsetzungen beispielsweise der Historie innerhalb eines Dialogs bedienen und immer dann nachfragen, wenn eine der einzelnen Feldeingaben die Schwelle unterschreitet.

Eine **Übertragung** der hier postulierten konfidenzabhängigen Rückholstrategie lässt sich auch auf weitere Anwendungsbereiche andenken. Zur Bestimmung der Schwellenhöhe sollte dabei, in Anlehnung an zwischenmenschliche Dialoge, der Kontext Beachtung finden. Je nach Aufgabenart, Nutzerkonfiguration oder Wortschatzähnlichkeit sollte die Schwelle unterschiedlich definiert werden. Als Transporteur von Emotionen könnte auch das Sprachsignal des Nutzers Anhaltspunkte zur Anpassung der Dialogstrategie bieten.

Grundsätzlich lautet die Empfehlung, dass je schwerwiegender die Folgen eine Missverständnis innerhalb eines Kontextes wären, desto höher sollte die Konfidenzschwelle gewählt werden. Es lassen sich sogar Kontexte antizipieren, in denen immer eine explizite Evidenz unabhängig der Erkennersicherheit erfolgen sollte. Beispielsweise bei einem möglichen Datenverlust, einer hohen Ladedauer oder monetären Transaktionen sollte immer eine Rückversicherung durch den Nutzer erfolgen. Auch wenn diese kontextspezifische Anpassung Aufwände im Dialogdesign erfordern, so zeigen die Ergebnisse, dass der Effizienzgewinn durch Zeit- und Dialogschritteinsparung beachtlich ist und die Nutzerzufriedenheit dadurch steigt.

Eine zusätzliche farbliche **Visualisierung** der Konfidenz wurde gewählt, um die inkonsistenten Reaktionen des dynamischen Systems zu erklären und das Grounding äquivalent zur menschlichen Mimik zu unterstützen. Hier konnte nachgewiesen werden, dass diese unterstützende Erklärung der Systemreaktionen bei dynamischen Dialogverläufen benötigt wird. Es zeigte sich, dass die Metapher zur Vermittlung der Systemsicherheit auch ohne Erleben der Anwendung die existierenden mentalen Modelle der Nutzer bediente. Die farbige Visualisierung konnte darüber hinaus positive Effekte bezüglich der Blickabwendung und Fahrleistung erzielen. So konnten die Versuchspersonen die farbige Gestaltung peripher gut wahrnehmen. Bei späteren Umsetzungen sollte allerdings die Farbgebung für das Mikrofon überdacht werden. So bildete das Grau kaum einen Unterschied zur Farbe des Hintergrunds.

Die Analyse der Ablenkungsparameter verdeutlichte, dass die **akustische Rückmeldung** der Aktionsankündigung alle relevanten Einheiten beinhalten sollte. Das Nennen des Namens, aber nicht des Nummerntyps führte zu anfänglichen Rückversicherungsblicken auf die textuelle Anzeige. Die Gestaltungsregel, dass die Visualisierungen zwar sinnvolle Ergänzungen sind, aber es auch gewährleistet sein muss, dass die Dialoge auch ohne Blickzuwendung erfolgreich abgeschlossen werden können, sollte zukünftig nicht verletzt werden.

6.2.4.3 Diskussion Methode

Es kann davon ausgegangen werden, dass aufgrund der kleinen **Stichprobe** und der daraus folgenden relativ geringen Teststärke sich weitere bedeutsame Effekte nicht signifikant zeigten (Bortz, 2005). Umso mehr kann die Relevanz der diskutierten Einflüsse betont werden.

Es wurde sich bezüglich des Dialogverhaltens für ein **Messwiederholungsdesign** entschieden. In der Regel bietet ein solches Vorgehen den Vorteil, dass Versuchspersonen unter mehreren Bedingungen getestet werden und so meist weniger Störvariablen vorhanden sind, was zu einer höheren Teststärke führt. Durch die Kontrastierung der Dialogverhaltensweisen kann der positive Effekt des dynamischen Systems jedoch auch leicht überschätzt werden.

Die Analysen zeigten darüber hinaus, dass die zeitliche Position der Bedingung einen Effekt auf das Ergebnis haben kann. Trotz einer vollständigen Ausbalancierung konnten Positionseffekte über alle VP nicht ausgeschlossen werden. Insbesondere Carry-over-Effekte, die den Einfluss der Erfahrung mit einer früheren experimentellen Bedingung beschreiben, konnten beobachtet werden. So reagierten Versuchspersonen beispielsweise empfindlich, wenn sie erst das statische System erlebten und die gewohnte Rückversicherung des Systems in der dynamischen Bedingung dann wegfiel. Innerhalb der Analysen musste die Störvariable des Positionseffekts immer kontrolliert werden.

In gewisser Weise kann das gemischte **Versuchsdesign** für die geringen Effekte des Feedbackverhaltens verantwortlich gemacht werden. Zur Kontrastierung der Effekte des flexiblen System Grounding Criterion wurde ein Design gewählt, bei dem jede Versuchsperson sowohl das statische, als auch das dynamische System erproben konnte, während jeweils nur eine Feedbackvariante erlebt wurde. Bei einem solchen gemischten Versuchsdesign werden die Effekte des Innersubjektfaktors stärker betont als die des Zwischensubjektfaktors. So werden bei dem angewandten Split-Plot-Verfahren unterschiedliche Fehlervarianzen für die Signifikanzprüfung der Inner- und Zwischensubjekteffekte zu Grunde gelegt. Damit werden zwar die Tests der Innersubjekt- und Interaktionseffekte optimiert, die Effektivität der Zwischensubjekteffekt-Tests aber eingeschränkt (Bradley & Russell, 1998). Somit kann es zu einer Unterschätzung der Effekte der Zwischensubjektfaktoren kommen. Für weiterführende Untersuchungen sollte daher in Erwägung gezogen werden, die Befunde in komplett randomisierter Vier-Felder-Form und mit einer a posteriori determinierten Stichprobengröße zu überprüfen.

Die **Skalenprüfung** der SUS zeigte eine ähnlich hohe interne Konsistenz wie sie u.a. von Sauro (2011) berichtet wird. Eine Faktorenanalyse zeigte, dass von einer Eindimensionalität des Fragebogens, wenn auch mit Einschränkungen, ausgegangen werden kann. Beides kann in Anhang 12.4.5 nachverfolgt werden.

Eine Skalenprüfung des selbsterstellten ITU-Fragebogens belegte, dass dieser nicht nur eine hohe interne Konsistenz und damit Reliabilität aufweist. Er zeigte im Rahmen einer Faktorenanalyse außerdem, dass er als eindimensional gelten kann (siehe Anhang 12.4.5). Eine weitere Prüfung dieser Skala mit größeren Stichproben oder anhand von externen Validitätskontrollen ist dennoch empfehlenswert.

Die Erhebung der subjektiven **Zufriedenheits- und Systembeurteilungen**, die bisher jeweils durch ein Einzelitem erfolgte, sollte ausgebaut werden. So könnte die Beurteilung um Aspekte des Joy-of-Use und der Intuitivität erweitert werden. Durch die Konstruktion ganzer Subskalen zu diesen Aspekten könnten Bewertungen über unterschiedlich gepolte Items differenzierter erfasst

und eine reliablere und validere Erhebung der Zufriedenheit und der Systembeurteilungen ermöglicht werden.

Wie bereits in Studie I erläutert, wirkte sich auch in dieser Studie die **Künstlichkeit der Laborsituation** in vielerlei Aspekten auf die Mensch-SDS-Interaktion aus (siehe Kapitel 6.1.4.2).

6.2.5 Zusammenfassung

Es konnte gezeigt werden, dass die Umsetzung des flexiblen Rückfrageverhaltens in Anlehnung an das Prinzip des Least Collaborative Effort (Clark & Wilkes-Gibbs, 1986) eine Erleichterung der Grounding-Prozesse ermöglichte und die nutzerseitigen Aufwandskosten minimierte. Ohne eine technische Verbesserung des Systems konnte allein über die Dialoggestaltung nach den Maximen der zwischenmenschlichen Kommunikation eine multidimensionale Steigerung der Gebrauchstauglichkeit erreicht werden. Insgesamt konnte belegt werden, dass sowohl die Effizienz, die Effektivität als auch die Zufriedenheit durch die Implementation einer adaptiven Interpretationsstrategie signifikant verbessert werden konnten.

Auch wenn ein Grundlevel an Effizienz, Effektivität und Zufriedenheit durch das Referenzsystem gewährleistet wurde, konnte es durch flexibles Rückfrageverhalten noch einmal deutlich gesteigert werden. Sowohl der objektive als auch subjektiv bewertete Zeitaufwand ließen sich durch das konfidenzbasierte Dialogverhalten verbessern. Bedienfehler im Dialogablauf, die sich u.a. durch verfrühtes Sprechen äußerten, wurden bei dynamischem Dialogverhalten zusätzlich reduziert. Auch die wahrgenommene Zuverlässigkeit der Befehlsausführung und die Zufriedenheit wurden verbessert.

Die Ergebnislage belegt, dass eine Übertragung zwischenmenschlicher Kommunikationsstrategien auf SDS als ein vielversprechender Ansatz gelten kann. Dabei hat auch inkonsistentes Systemverhalten keinen negativen Einfluss auf die Systembewertung, wenn es für den Nutzer ausreichend transparent geschieht. Anhand der Befunde konnte dokumentiert werden, dass das Prinzip des Least Collaborative Effort (ebd.) auch im Rahmen der Mensch-Maschine-Dialoge Gültigkeit besitzt und die Maxime *Fasse dich kurz.* (Grice, 1975; Kapitel 3.1.2.1) bedeutsamer ist als die Systemkonsistenz.

Die folgenden Gestaltungsempfehlungen lassen sich aus der Untersuchung ableiten:

- Die **Erkennerkonfidenz** kann als hinreichender Parameter zur Flexibilisierung des Dialogverhaltens gelten. Dennoch sollten kontextabhängige Definitionen des Schwellenwertes erfolgen.

- Die Analysen zeigten, dass die **liberalste Schwellendefinition** von 40%[31] zu der größten Effizienz- und Effektivitätssteigerung führte.
- Das Rückfrageverhalten sollte von Beginn des Dialoges an dynamisch gestaltet werden. Eine **Anpassung** im Verlauf kann die nutzerseitige Akzeptanz schmälern.
- Im Rahmen der finalen Aktionsankündigung sollten **alle Informationseinheiten** akustisch dargeboten werden.
- Die farbige Vermittlung der **Systemsicherheit** über die Ampelmetapher im Rahmen des Zustandsfeedback verhilft dem Nutzer, ein mentales Modell des Dialogverlaufs aufzubauen.

Nachdem im Rahmen der Studie II die Bedeutsamkeit der Effizienz im Dialog mit maschinellen Gesprächspartnern nachgewiesen werden konnte, soll Studie III die strategischen Aspekte der Mensch-Dialogsysteminteraktion beleuchten.

[31] Hierbei handelt es sich um einen bedingt generalisierbarer Schwellenwert.

6.3 Studie III: Systemseitiges Alignment

Bezugnehmend auf das dritte Element des Groundings, soll an dieser Stelle die Angleichung der Systemausgaben an die Nutzereingabe untersucht werden. Zur Generierung der Systemantwort passte sich das System lexikalisch an die Nutzereingabe an. Dabei galt es zu erforschen, ob diese erhöhte Systemadaptivität Einfluss auf die Systembewertung und die nutzerseitig wahrgenommene Systemkompetenz hat. Das angleichende System wird dabei gegen ein System verglichen, welches bewusst andere (lexikalische) Formulierungen als der Nutzer verwendet und gegen ein System mit gleichbleibender Wortwahl. Die im folgenden Abschnitt aufgeführten Annahmen sollten dabei getestet werden.

6.3.1 Hypothesen

In Abgrenzung zu Porzels (2006) Erhebung soll explizit keine Verbesserung der Effizienz des Systems mit der Angleichung einhergehen, um die Forschungsfrage zu adressieren, ob linguistisches Alignment auch ohne eine objektive Zeiteinsparung zu besseren Systembeurteilungen führt. Dafür sollten folgende Annahmen gelten:

1. **Zwischen den drei Dialogbedingungen zeigen sich keine Effizienzunterschiede.**

Abgeleitet aus den in Kapitel 3 vorgestellten Untersuchungen des zwischenmenschlichen Dialogverhaltens wird davon ausgegangen, dass das lexikalische Alignment des SDS den Aufgabenerfolg und die nutzerseitige Zuverlässigkeitsbeurteilung steigern kann. So wird angenommen, dass bei einer systemseitigen Angleichung an die Wortwahl des Nutzers das gegenseitige Verständnis schneller erreicht und unnötige Korrekturen (vgl. Verkehrsfunkbeispiel, Hassel, 2006) vermieden werden können. Auch im Bereich der Fehlbedienungen kann aufgrund von einer Reduktion des TTE-Fehlers durch die Angleichung ausgegangen werden. So konnte Nenkova (2008) in Untersuchungen von zwischenmenschlichen Dialogen belegen, dass die lexikalische Angleichung zu weniger Unterbrechungen führte.

2. **Die lexikalische Angleichung durch das System erhöht die Effektivität der Mensch-Maschine-Interaktion.**

Abgeleitet aus den Erkenntnissen von Metzinger und Brennan (2003) kann davon ausgegangen werden, dass Nutzer sogar mit der Erwartungshaltung in einen Dialog kommen, dass bereits akzeptierte Begriffe verwendet werden und daher Systeme präferieren, die diesem mentalen Modell gerecht werden.

3. **Es zeigt sich eine höhere nutzerseitige Zufriedenheit und Präferenz für ein angleichendes System.**

Aus den letzten beiden Annahmen ergibt sich die Hypothese, dass ein angleichendes System zwei der drei Gebrauchstauglichkeits-Kriterien nach DIN EN ISO 9241-11 (1998) besser erfüllt, als ein nicht-angleichendes System.

4. **Die lexikalische Angleichung erhöht die Gebrauchstauglichkeit des SDS.**

In diesem Zusammenhang wird angenommen, dass das Alignment ebenfalls die verhaltensbasierte Bereitschaft Sprachbedienung zu benutzen erhöhen sollte.

5. **Die lexikalische Angleichung erhöht die Nutzungsbereitschaft des SDS.**

Betrachtet man die strategischen Aspekte des Alignments, so spricht die Befundlage dafür, dass (menschliche) Sprecher, die eine Angleichung des Wortschatzes zeigten, positiver beurteilt werden (u.a. Bradac et al., 1988), wohingegen die Wahl von alternativen Äußerungen als korrigierend empfunden wird und negative Affekte auslösen kann (Branigan et al., 2010). Durch eine Angleichung sollte das Sprachdialogsystem nicht nur sympathischer und höflicher, sondern auch menschenähnlicher erscheinen. Ein System, welches bewusst andere Formulierungen verwendet, sollte demnach als weniger höflich und sympathisch eingestuft werden. So konnte Nenkova (2008) belegen, dass eine lexikalische Angleichung ein guter Indikator für eine wahrgenommene Natürlichkeit der Unterhaltung ist.

6. **Die lexikalische Angleichung erhöht die Sympathie- und Natürlichkeitsbeurteilungen.**

Erneut würde ein adaptives System durch die Angleichung an die Nutzereingabe in gewisser Weise inkonsistente Reaktionen zeigen. Wie bereits in Studie II wird jedoch davon ausgegangen, dass diese Inkonsistenz nicht wahrgenommen wird und zu keinen negativen Effekten führt.

7. **Die adaptiven Systemreaktionen des angleichenden Systems führen zu keiner negativen Konsistenzbeurteilung.**

In Anlehnung an die Befunde von Porzel (2006) wird davon ausgegangen, dass die sprachliche Kompetenz des adaptiven Systems am besten bewertet wird.

8. **Die lexikalische Angleichung erhöht die wahrgenommene Sprachkompetenz des Systems.**

Durch die unter 3.2.3 erläuterten Primingprozesse kann das Alignment auch zu einer Erleichterung der Produktion und des Verstehens im Dialog führen. Gelten die unvermittelten Prozesse des Alignments auch in der Mensch-Maschine-Kommunikation, so kann davon ausgegangen

werden, dass ein SDS, welches eine Anpassung an die Eingabe zeigt, den Verarbeitungsaufwand auf Seiten des Nutzers reduzieren kann. Im Bereich des Groundings könnten also Kosten hinsichtlich der Rezeption und des Verständnisses gespart werden und der kognitive Aufwand der Sprachbedienung weiter reduziert werden.

9. **Die lexikalische Angleichung erleichtert die Rezeption und reduziert so die mentale Beanspruchung der Dialogführung. Die Ablenkung von der Fahraufgabe ist bei der Interaktion mit dem angleichenden System am geringsten.**

Explorativ sollen in Anlehnung an Porzel (2006), der einen größeren Nutzwert des Alignments in der Nutzergruppe der Novizen finden konnte, Unterschiede zwischen Nutzern mit **Erfahrung** im Bereich Sprachdialogsysteme und Novizen dokumentiert werden.

Das folgende Kapitel soll Einblicke in den methodischen Kontext der dritten Studie geben, deren Verlauf im folgenden Absatz dargestellt wird.

6.3.2 Methode

In Abgrenzung zu den beiden ersten Untersuchungen findet zunächst eine Beschreibung der Fahr- und Sprachbedienaufgaben statt. Des Weiteren werden kurz die Erhebungsinstrumente eingeführt und begründet. Anschließend wird das Untersuchungsdesign, die Stichprobenzusammensetzung und der Versuchsablauf erläutert.

6.3.2.1 Fahraufgabe

Die Untersuchung fand in einer Sitzkiste in den Räumlichkeiten der Volkswagen Konzernforschung statt. Diese Sitzkiste stellt eine rudimentäre Nachbildung eines Fahrzeugs mit Automatikgetriebe dar und besteht aus Fahrersitz, Lenkrad und Pedalerie. Sie ist in Längs- und Querführung unbeweglich, so dass Fahrbewegungen nur optisch simuliert werden. Die Streckensimulation erfolgt auf einem 17'' PC-Monitor, der vor der Sitzkiste aufgebaut war (Abbildung 47). Die Lautsprecher des Monitors wurden genutzt, um Fahrgeräusche auszugeben.

Als primäre Fahraufgabe wurde auf die Software LCT (Lane Change Task) zurückgegriffen. Der LCT wurde als Werkzeug zur Bewertung der Fahrerablenkung entwickelt und stellt eine der echten Fahraufgabe äquivalente, kognitive Belastung dar (Mattes 2003, Harbluk et al. 2007). Die Software hat sich inzwischen zum Industriestandard zur Messung der Fahrablenkung etabliert und bietet die Möglichkeit zur realitätsnahen und effizienten Testung von Fahrerinformationssystemen (ISO, 2011).

Abbildung 47: Versuchsaufbau Studie III

Der LCT fand bei der vorliegenden Studie in leicht abgeänderter Form Anwendung. So in der Simulation ein Vorderfahrzeug ergänzt, um neben der Spurhaltung und der Geschwindigkeit auch den Abstand zum Vorderfahrzeug zu erfassen. Die Versuchspersonen wurden für alle Fahrten instruiert, dem Vorderfahrzeug mit möglichst konstanter Geschwindigkeit von 60 km/h und einem Abstand von 50 Metern zu folgen. Ferner wurde auf die Aufforderung zum Spurwechsel verzichtet. Die Fahraufgabe der Probanden bestand demnach im Wesentlichen darin, die Längs- und Querführung mittels eines Logitech Lenkrads und angeschlossener Pedalerie auf der mittleren Spur eines dreispurigen Autobahnszenarios zu kontrollieren.

Entsprechend der ersten beiden Studien, kann demnach auch im vorliegenden Fall von einer Fahraufgabe mit geringer Schwierigkeitsstufe ausgegangen werden.

6.3.2.2 Sprachdialogsystem

Der Untersuchung lag erneut das in Kapitel 6.1.2.2 beschriebene Dialogsystem zugrunde. Der PTT-Knopf zur Aktivierung des SDS befand sich in dieser Untersuchung nicht am Lenkrad der Sitzkiste, sondern lag auf einer Ablage neben dem Lenkrad. Die Interaktion mit dem SDS erfolgte mittels eines an der Sitzkiste montierten Mikrofons und zwei Standard-PC Lautsprechern, die hinter der Versuchsperson angebracht waren.

6.3.2.2.1 Dialogbedingung

Zur Evaluation der Effekte des Alignments wurden drei Dialogvarianten, basierend auf verschiedenen Rückmeldekonzepten (angleichend, konstant, abweichend), konzipiert. Bei gleicher Semantik unterschieden sich die Systemrückmeldungen je nach Dialogbedingung lexikalisch. Dabei unterschied sich die Wortwahl des Systems, wenn es nach dem Kontakt fragte, sich den Auftrag explizit bestätigen lies oder den Aufbau des Telefonanrufs ankündigte. Durch die Flexibilisierung

der Frage nach dem Kontakt, hätte auch bei einer schrittweisen Eingabe die Manipulation erlebt werden können.

Das Dialogverhalten, wie es aktuelle Sprachdialogsysteme zeigen, wurde durch konstante Systemrückmeldungen realisiert. Wie in Tabelle 32 dargestellt, reagierte das System auf eine Nutzereingabe immer mit der gleichen Wortwahl (im Folgenden als **off-System** bezeichnet). Objektiv wies dieses off-System den geringsten Wortschatz auf, da es immer die gleichen Begrifflichkeiten benutzte.

Tabelle 32: Systemausgaben in der off-Bedingung

Nutzereingabe	**Systemreaktion (explizite Bestätigung)**	**Systemreaktion (Handlungsankündigung)**
anrufen	"Möchten Sie Helga Otto anrufen?"	"Es wird gewählt."
wählen	"Möchten Sie Helga Otto anrufen?"	"Es wird gewählt."
Verbindung herstellen	"Möchten Sie Helga Otto anrufen?"	"Es wird gewählt."
telefonieren	"Möchten Sie Helga Otto anrufen?"	"Es wird gewählt."
sprechen	"Möchten Sie Helga Otto anrufen?"	"Es wird gewählt."

Darüber hinaus wurde ein System umgesetzt, das den Nutzerinput lexikalisch und auch teilweise syntaktisch[32] spiegelte. Das sogenannte **on-System** zeigte ein wortbasiertes Alignment und damit eine hohe Flexibilität bei den sprachlichen Rückmeldungen. Auch bei dieser adaptiven Schnittstelle handelt es sich um ein generisches Benutzermodell (BM), welches sich an jeden Nutzer gleichermaßen anpasst. Es handelt sich weiterhin um ein dynamisches BM, in welches noch während der Interaktion Informationen zur Adaption einfließen. Ein typischer Dialogablauf mit dem adaptiven System ist im Folgenden dargestellt.

[32] Eine teilweise syntaktische Spiegelung konnte dadurch erreicht werden, dass dem Nutzer die Kommandos, die er einsprechen sollte genau vorgegeben wurden. Somit konnte auch der Eindruck einer hohe Passung der Satzstruktur erreicht werden.

Nutzer	Anrufen.
System	Wen möchten sie anrufen?
Nutzer	Ich möchte Helga Otto anrufen.
System	Möchten Sie Helga Otto anrufen?
Nutzer	Ja.
System	Es wird angerufen.

Da der technische Aufwand bei der Implementation des lexikalischen Alignments als hoch bezeichnet werden kann, wurde die systemseitige Angleichung nur für das Formular zum Starten eines Telefonanrufs realisiert. Für insgesamt fünf Kommandos wurde eine lexikalische Angleichung des Systems innerhalb von drei Dialogzuständen umgesetzt.

In der folgenden Tabelle 33 können die möglichen angeglichenen Antworten des Systems auf die Nutzereingaben eingesehen werden.

Tabelle 33: Systemausgaben in der on-Bedingung

Nutzereingabe	**Systemreaktion (Teilnehmer missing)**	**Systemreaktion (explizite Bestätigung)**	**Systemreaktion (Handlungsankündigung)**
anrufen	"Wen möchten Sie anrufen?"	"Möchten Sie Helga Otto anrufen?"	"Es wird angerufen."
wählen	"Wessen Nummer möchten Sie wählen?"	"Die Nummer von Helga Otto wählen?"	"Es wird gewählt."
Verbindung herstellen	"Mit wem wollen Sie verbunden werden?"	"Verbindung mit Helga Otto herstellen?"	"Verbindung wird hergestellt."
telefonieren	"Mit wem wollen Sie telefonieren?"	"Mit Helga Otto telefonieren?"	"Telefonat wird gestartet."
sprechen	"Mit wem wollen Sie sprechen?"	"Möchten Sie mit Helga Otto sprechen?"	"Gespräch wird hergestellt."

Da das on-System per Definition somit einen höheren Wortschatz als das konstante off-System aufweist, sollte ein weiteres System geschaffen werden, das zwar objektiv die gleiche Sprachkompetenz wie das on-System besaß, aber keine Angleichung zeigte. Das sogenannte **wrong-System** verknüpfte jede Nutzereingabe mit einer anderen Begrifflichkeit und verwendete diese für die Generierung der Systemprompts. In den meisten Fällen unterschied sich die Handlungsankündigung von der Bestätigungsaufforderung hinsichtlich der Wortwahl (Tabelle 34).

Tabelle 34: Systemausgaben in der wrong-Bedingung

Nutzereingabe	Systemreaktion (explizite Bestätigung)	Systemreaktion (Handlungsankündigung)
anrufen	"Die Nummer von Helga Otto wählen?"	"Es wird gewählt."
wählen	"Verbindung mit Helga Otto herstellen?"	"Verbindung wird hergestellt."
Verbindung herstellen	"Mit Helga Otto telefonieren?"	"Es wird angerufen."
telefonieren	"Möchten Sie mit Helga Otto sprechen?"	"Es wird gewählt."
sprechen	"Möchten Sie Helga Otto anrufen?"	"Telefonat wird gestartet."

Das wrong-System sollte auch dem Nachweis negativer Effekte bei bewusster Verwendung anderer Begrifflichkeiten dienen.

Alle drei Systeme zeigten weder eine ergänzende Visualisierung, noch eine flexible Anpassung des Rückfrageverhaltens. Durch die statische Dialogstruktur konnte erreicht werden, dass jede Dialogvariante mindestens bei zwei Systemausgaben durch die Versuchsperson erlebt werden konnte.

6.3.2.2.2 Kommandos

Erneut fanden auf der Basis von Erfahrungswerten Kommandos Anwendung, die in den vorangegangenen Untersuchungen häufig erfolgreich abgeschlossen und mit ausreichender Güte erkannt wurden.

Entsprechend dieser Kriterien wurden sieben Kommandos zum Starten eines Telefonanrufs ausgewählt. Für jedes Kommando war eine Mindestanzahl von zwei Sprecherwechseln nötig. Eine Übersicht der Kommandos kann Anhang 12.5.3 entnommen werden.

6.3.2.3 Versuchsplan

Um die Effekte der verschiedenen Rückmeldestrategien zu kontrastieren, wurde ein Messwiederholungsdesign gewählt. Jeder Proband erlebte jedes der Systeme in randomisierter Reihenfolge. Die Systeme unterschieden sich danach, ob sich das SDS an die Nutzereingabe anglich (on), immer konstant reagierte (off) oder bewusst andere Formulierungen verwendete (wrong). Der Nutzerstudie lag somit ein einfaktorieller 3-Felder Versuchsplan (Alignment: on vs. off vs. wrong)

mit Messwiederholung zugrunde. Um den Effekt der zeitlichen Position der Dialogbedingung zu reduzieren, fand ein vollständiges Ausbalancieren statt.

6.3.2.4 Erhebungsinstrumente

Zur Erfassung der globalen Gebrauchstauglichkeit fand erneut die **System Usability Scale** (SUS) Anwendung (Kap. 6.1.2.4). Angeschlossen an die Items der SUS fand eine subjektive Einschätzung der empfundenen Ablenkung über ein Item *Ich fühlte mich von der Fahraufgabe abgelenkt.* statt.

Den objektiven Maßen lagen auch in dieser Studie Beobachtungen durch geschulte Versuchsleiter zugrunde. Demzufolge wurde die Effektivität durch die objektiven **Protokollmaße** erfolgreiche Dialogabschlüsse, Fehlerkennungen, Fehlbedienungen (z. B. zu frühe Spracheingabe) und subjektive Zuverlässigkeitsbeurteilungen operationalisiert. Die Effizienz wurde über die Anzahl der Sprecherwechsel, die durchschnittliche Dialoglänge und subjektive Einschätzungen der Aufwandsangemessenheit erfasst. In der on-Bedingung wurde weiterhin notiert, wenn durch eine Fehlerkennung keine Angleichung durch das System gezeigt wurde. Dies war immer dann der Fall, wenn die Formulierung (z. B. anrufen oder wählen) oder der Syntax nicht korrekt in die Aktion oder Bestätigungsanfrage aufgenommen wurden.

Während der Referenz- und Aufgabenfahrten wurden die **Fahrdaten** über die LCT-Software erfasst. Diese bot durch das Setzen von Markierungen auch die Möglichkeit, die Interaktionsdauer zu bestimmen und damit ein objektives Maß für die Dialoglänge zu erfassen.

Zur differenzierten Erfassung der nutzerseitigen Einschätzung der **sprachlichen Kompetenz** des Systems (Wortschatzgröße, Wortgewandtheit, Fähigkeit des Systems Sprache zu verstehen, Grammatikkenntnisse), der Beurteilung der **Natürlichkeit** (Intelligenz, Sympathie, Höflichkeit, Menschenähnlich) und der **Zufriedenheit** fanden primär selbstkonstruierte Items der Kategorienunterteilungsskalen Anwendung (siehe Anhang 12.5.4).

Ebenso wurde der Intention to Use Fragebogen aus Studie II zur Erfassung der **Nutzungsbereitschaft** verwendet (siehe Kap. 6.1.2.4). Erneut wurde er vor und nach den Aufgabenfahrten erhoben und soll somit eine mögliche individuelle Veränderung der Nutzungsbereitschaft durch die Interaktion mit den verschiedenen Dialogbedingungen abbilden. Der ITU wurde dahingehend verändert, dass seine Items nun spezifischer formuliert wurden. Während bei der Prä-Messung noch nach Sprachbedienung allgemein gefragt wurde, zielten die Fragen nach den Aufgabenfahrten im Wesentlichen auf das erlebte System ab (siehe Tabelle 35).

Tabelle 35: Items des ITU-Fragebogens Studie III

ITU Post-Messung	**--**	**-**	**0**	**+**	**++**
Ich halte das eben benutze Sprachbediensystem für eine sinnvolle Alternative zur Tasteneingabe.	O	O	O	O	O
Diese Sprachbedienung hat mich angestrengt.	O	O	O	O	O
Ich kann mir gut vorstellen, diese Sprachbedienung zu nutzen.	O	O	O	O	O
Ich finde diese Sprachbedienung fehleranfällig.	O	O	O	O	O
Ich würde diese Sprachbedienung als Zusatzausstattung wählen.	O	O	O	O	O
Ich finde diese Sprachbedienung überflüssig.	O	O	O	O	O

Im Anschluss an die Befragung wurden die **Tops und Flops** der jeweiligen Dialogbedingung offen erfragt. Der letzte Durchgang beinhaltete noch eine Präferenzbeurteilung. Eine Übersicht aller Erhebungsinstrumente liefert der Versuchsablauf in Anhang 12.5.4.

6.3.2.5 Stichprobe

Am Versuch nahmen ausschließlich Mitarbeiter der Volkswagen AG teil. Für die Versuchsteilnahme wurde der Besitz eines Führerscheins sowie Deutsch als Muttersprache vorausgesetzt. Entsprechend dieser Kriterien wurden 25 Personen aus der Datenbank des betriebsinternen Probandenpools rekrutiert. Fünf Personen konnten aufgrund von technischen Problemen nicht alle drei Dialogbedingungen erleben und wurden bei den Varianzanalysen mit Messwiederholung ausgeschlossen, sodass insgesamt eine Stichprobe von N= 20 Personen zur Verfügung stand.

Die Altersspanne betrug 23 bis 59 Jahre (*MW*= 39.0 Jahre; *SD*= 8.4 Jahre). Mit 12 weiblichen und 8 männlichen Probanden wurde eine annähernde Gleichverteilung der Geschlechter erreicht. 12 Versuchspersonen gaben an, ein technisches Gerät zu besitzen, welches über Sprachbedienung gesteuert werden kann (Navigationsgerät, Mobiltelefon, Freisprecheinrichtung im Auto), während die verbleibenden 8 Probanden kaum Erfahrung mit Sprachbedienung hatten oder nur inzidentelle Nutzung (Diktiergerät, Auskunftssysteme) berichteten. Dabei stand der Erfahrungsgrad nicht in Verbindung mit dem Geschlecht der Probanden.

6.3.2.6 Durchführung

Vor Studienbeginn wurden die Versuchsteilnehmer informiert, dass es sich bei der geplanten Untersuchung um einen Sitzkistenversuch zur Bewertung eines neuartigen Sprachdialogsystems zur Telefonbedienung handelte und wurden über die Aufzeichnung der Fahrdaten in Kenntnis

gesetzt. Es folgte die soziodemografische Befragung, in der auch Erfahrungswerte und die Basisnutzungsbereitschaft erfragt wurden. Im Anschluss daran wurden alle Teilnehmer standardisiert für die Fahraufgabe instruiert.

Innerhalb der Instruktion in das Sprachdialogsystem wurden die Teilnehmer explizit darauf hingewiesen, dass sie drei Systeme erleben werden, die sich hinsichtlich ihrer Wortwahl unterscheiden. Die VP wurden darüber informiert, dass ein exaktes Nachsprechen der Kommandos erforderlich sei, es erfolgte aber keine explizite Einführung in das Alignment des on-Systems. Allen Probanden wurde im Stand das Beispielkommando *Julia Sommer zu Hause anrufen.* vorgeführt, welches einen ersten Einblick in das Rückmeldekonzept geben sollte. Jedes System wurde anhand des gleichen Beispielkommandos eingeführt. Die standardisierte Instruktion kann dem Anhang 12.5.2 entnommen werden.

Jeder Teilnehmer absolvierte eine Eingewöhnungsfahrt, drei Aufgabenfahrten, sowie eine Referenzfahrt vor und nach den Aufgabenfahrten. Während der Aufgabenfahrten wurden sieben typische Sprachkommandos zum Vergleich der drei Dialoglösungen ausgeführt.

Die Systembeurteilungen der SUS- und ITU-Fragebögen, erfolgten nach Beendigung der Aufgabenfahrten. Die Erhebung dauerte durchschnittlich 45 Minuten. Ihr schematischer Ablauf kann dem Anhang 12.5.1 entnommen werden.

6.3.3 Ergebnisse

Im Rahmen dieses Abschnitts werden bedeutsame Effekte der Dialoggestaltung auf die erhobenen Variablen berichtet. Gemäß den Forschungshypothesen wird durch Angleichung der Formulierung an den Nutzer ein positiver Effekt auf die Gebrauchstauglichkeit erwartet. Daher liegt der Fokus der Analysen auf dem Dialogverhalten. Zur Auswertung des einfaktoriellen Versuchsdesigns fanden (M)ANOVAs mit Messwiederholung statt.

Zur Prüfung der Voraussetzungen sei auf Kapitel 6.1.2.7 verwiesen. Im Falle einer Verletzung der Sphärizitätsannahme fanden entsprechende Korrekturen Anwendung. Durch den quasi-experimentell erfassten Faktor Erfahrung ergaben sich auf jenem Zwischensubjektfaktor keine identischen Zellenbesetzungen, was bei der Interpretation der berichteten Ergebnisse Beachtung finden sollte.

Bevor die Effekte des Dialogverhaltens und der Feedbackdarstellung auf die Usability-Aspekte berichtet werden, soll durch eine Betrachtung der Anzahl angeglichener Äußerungen kurz die **Manipulation** des Dialogverhaltens überprüft werden. Über alle Versuchspersonen hinweg konnten in der on-Bedingung nur drei Kommandos beobachtet werden, bei denen das System kein

Alignment zeigte. Somit konnten 97.9% der Nutzereingaben angeglichen werden. Ein fehlendes Alignment ging in den meisten Fällen mit einer Fehlerkennung einher.

Es wurde ebenfalls durch eine qualitative Analyse der Tops/Flops-Antworten erfasst, wie viele Probanden sich der Angleichung des Systems innerhalb der on-Bedingung bewusst waren. Nur eine Person bemerkte die Angleichung der Wortwahl bei der Systembeurteilung. 20% der Probanden bemerkten die Größe bzw. Flexibilität des Wortschatzes in der on-Bedingung. Letzteres soll im Folgenden als Alignment-Awareness bezeichnet werden.

6.3.3.1 Effizienz, Gebrauchstauglichkeit und Nutzungsbereitschaft

In Anlehnung an die Studie II wurden zur Bestimmung der **Effizienz** die durchschnittliche Dialoglänge und die Anzahl der Turn Takes als objektive Maße gewählt. Ebenso wurden die Einschätzungen der Dialoglänge innerhalb der KU-Ratings als subjektives Maß mit in die multivariate Varianzanalyse mit Messwiederholung aufgenommen.

Das Dialogverhalten zeigte auf das Konstrukt der Effizienz keinen signifikanten Einfluss (F[6,50]< 1; p=.76). So belief sich die Bearbeitungszeit der Aufgaben in allen drei Bedingungen auf ca. 16 Sekunden pro Kommando und es wurden durchschnittlich zwei Sprecherwechsel benötigt. Die Angemessenheit der Dialoglänge wurde bei jedem Rückmeldekonzept als *gut* eingeschätzt.

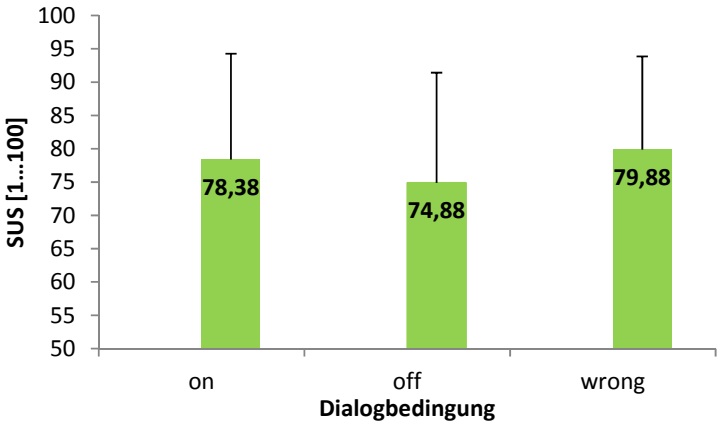

Abbildung 48: SUS-Scores (0 = Minimalwert 100 = Maximalwert) in Abhängigkeit der drei Dialogbedingungen

Zur Analyse des Einflusses der Dialogbedingungen auf die Usability-Einschätzung wurde der Gesamtscore des SUS in eine univariate Varianzanalyse mit Messwiederholung einbezogen. Jener SUS-Score lag in allen drei Versuchsbedingungen über 70 Punkten (siehe Abbildung 48) und damit über dem Durchschnittswert den Bangor et al. (2008) und Sauro (2011) ermittelten. Es

zeigten sich allerdings keine signifikanten Differenzen zwischen den Dialogbedingungen (F[1.5,29]= 1.6; p= .22).

In Anlehnung an die ersten beiden Untersuchungen wurden zur Erfassung der **Effektivität** neben den erfolgreichen Dialogabschlüssen, auch die Fehlbedienungen (zu frühe Spracheingabe) und Systemfehler (Fehlerkennung) analysiert. Es erfolgte in dieser Untersuchung keine Relativierung an der Summe der Turn Takes, da diese sich nicht zwischen den Dialogbedingungen unterschieden (Kap. 6.3.3.1). Zusätzlich wurde die subjektive Effektivität über das KU-Rating der wahrgenommenen Zuverlässigkeit der Kommandoausführung erfragt. Für das Effektivitätskriterium zeigte sich kein signifikanter Effekt des Innersubjektfaktors Dialogverhalten (F[8,68]= 1.0; p= .46). Insgesamt wurde die Zuverlässigkeit des Systems über alle Bedingungen hinweg als *gut* bewertet. Die Anzahl der erfolgreichen Dialogabschlüsse näherte sich in allen Dialogbedingungen 7 an, was einer Abschlussquote von 100% entsprach. Angeschlossene, heuristische Kontrastanalysen zeigten allerdings, dass das SDS in der on-Bedingung tendenziell zuverlässiger eingestuft wurde als das wrong-System (F[1,19]= 3.4; p= .08; η^2= .15).Bezog man den Zwischensubjektfaktor **Erfahrung** in die multivariate Varianzanalyse mit Messwiederholung mit ein, so zeigte sich ein signifikanter Haupteffekt des Expertiselevels auf die Zuverlässigkeitsbeurteilung (F[1,18]= 8.7; p= .01; η^2= .33). So schätzten Probanden ohne Erfahrung alle Dialogbedingungen als unzuverlässiger ein, als Probanden mit Erfahrung.

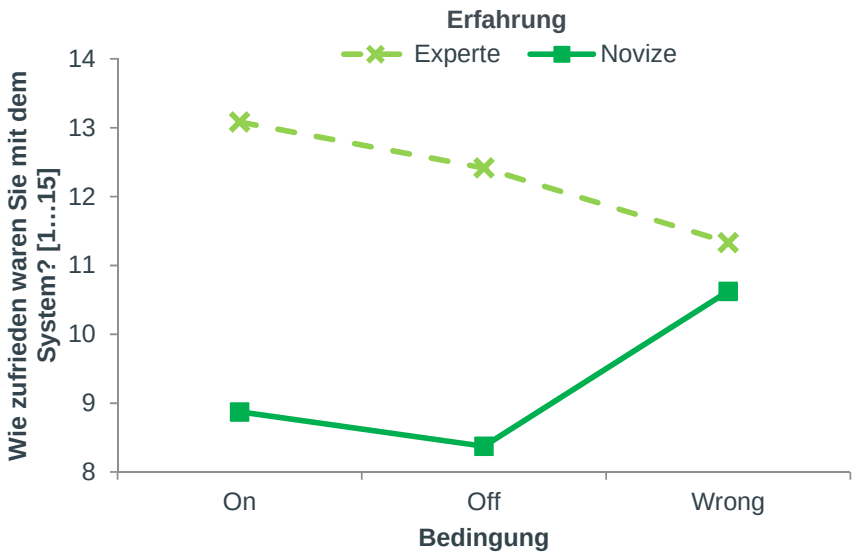

Abbildung 49: Zufriedenheitsbeurteilung (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Dialogbedingung und der Erfahrung

Betrachtet man die subjektive **Zufriedenheitsbeurteilung** innerhalb einer univariaten Varianzanalyse mit Messwiederholung, so konnte kein Haupteffekt des Innersubjektfaktors Dialogverhalten auf das Konstrukt Zufriedenheit festgestellt werden (F[1.2,23.5]< 1; p= .62). Betrachtet man den Zwischensubjektfaktor Erfahrung, so zeigte sich ein signifikanter Interaktionseffekt dieses Faktors mit dem Dialogverhalten (F[1.2,22.3]= 3.9; p=.05; η^2= .18). Während die Experten mit der on-Bedingung zufriedener waren als mit der wrong-Bedingung, verhielt es sich bei den Novizen genau umgekehrt. Darüber hinaus zeigte sich ein signifikanter Haupteffekt des Faktors Erfahrung, wobei die Nutzer mit Expertise insgesamt zufriedener mit allen Bedingungen waren, als Nutzer ohne Erfahrung (F[1,18]= 12.3; p< .01; η^2= .41) (siehe Abbildung 49).

Die subjektive **Präferenz** der Probanden für ein Dialogverhalten zeigte eine stark gleichverteilte Zuordnung (siehe Abbildung 50). Die Mehrzahl der Probanden hatte sich für das konstant reagierende System aus der off-Bedingung entschieden (40%), wohingegen 30% das on- und 30% das wrong-System wählten.

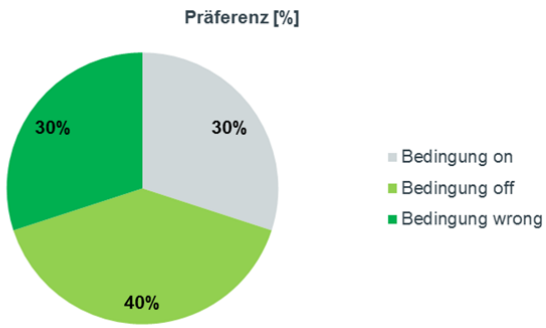

Abbildung 50: Prozentuale Verteilung der subjektiven Präferenz der drei Dialogbedingungen

Dabei zeigte sich ein Trend der Alignment-Awareness. So wurde das on-System tendenziell häufiger gewählt, wenn sich die Versuchspersonen der Angleichung bewusst waren (χ^2 [1, N= 20]= 4.8, p= .061)[33].

Zur Bestimmung der **Nutzungsbereitschaft** fand erneut der selbst erstellte Fragebogen ITU zu vier Messzeitpunkten Anwendung. Zur Analyse des Einflusses der Dialogbedingungen auf die Nutzungsbereitschaft wurden die sechs Facetten des ITU in eine multivariate Varianzanalyse mit Messwiederholung einbezogen. Die Dialogbedingung, die in diesem Fall auch die Prä-

[33] Nach Dichotomisierung der Präferenz wurde Fishers exakter Test durchgeführt.

Interaktions-Messung beinhaltet, zeigte auf das Konstrukt der Nutzungsbereitschaft einen tendenziell signifikanten Einfluss (F[18,143]= 1.8; p= .03; η^2= .18). Dieser ging maßgeblich auf Gruppenunterschiede in dem Bereich Anstrengung (*Sprachbedienung strengt mich an*.) und Fehleranfälligkeit des Systems (*Sprachbedienung ist zu fehleranfällig.*) zurück.

Bei beiden Items zeigte sich, dass nach den Systeminteraktionen eine positivere Bewertung der Nutzungsbereitschaft vorgenommen wurde (Anstrengung: F[3, 54]= 5.8; p< .01; η^2= .25; Fehleranfälligkeit: F[3,54]= 4.9 p= .01; η^2= .21). Der positive Nutzungseffekt der Anstrengungsdimensionen konnte unabhängig von dem Dialogverhalten (on, off oder wrong) beobachtet werden. Wohingegen der Nutzungseffekt bei der Beurteilung der Fehleranfälligkeit nur bei der on- und wrong-Bedingung positive Effekte gegenüber der Prämessung zeigte (on: F[1,18]= 9.5; p= .01; η^2= .35; wrong: F[1,18]= 9.5; p= .01; η^2= .35). Die off-Bedingung führte zu keiner signifikanten Reduktion der Fehleranfälligkeitsratings verglichen zur Prämessung (F[1,18]= 2.5; p= .12).

Bezog man den Faktor Erfahrung mit in die MANOVA ein, zeigten sich Haupteffekte auf den Dimensionen *Ich kann mir gut vorstellen diese Sprachbedienung zu nutzen.*, *Ich halte das eben benutzte Sprachbediensystem für eine sinnvolle Alternative zu Tasteneingabe.*, *Ich würde diese Sprachbedienung als Zusatzausstattung wählen.* und *Ich finde diese Sprachbedienung überflüssig.* (siehe Tabelle 36).

Tabelle 36: Effekte Erfahrung auf Nutzungsbereitschaft

	Effekt Erfahrung		
	$F_{(1,17)}$	**p**	**η^2**
…Sprachbedienung zu nutzen.	**3.7**	**.01**	**.36**
…sinnvolle Alternative zu Tasteneingabe.	**2.8**	**.01**	**.32**
…als Zusatzausstattung wählen.	**7.5**	**.01**	**.38**
…Sprachbedienung überflüssig.	**3.6**	**.02**	**.30**
…anstrengend.	0.4	.34	
…zu fehleranfällig.	1.0	.10	

Nutzer mit Expertise in der Nutzung von Sprachdialogsystemen schätzten diese auf allen Dimensionen besser ein, als Probanden, die nur geringe Erfahrung mit SDS aufwiesen. Darüber hinaus zeigte sich ein tendenzieller Interaktionseffekt der Dialogbedingungen mit der Erfahrung der Probanden auf den Dimensionen *Ich finde diese Sprachbedienung überflüssig.* (F[1,17]= 3.9; p= .07; η^2= .19). Bei der Einschätzung, ob Sprachbedienung eine überflüssige Ausstattung darstellt, zeig-

te sich, dass Experten dieser Aussage in der wrong-Bedingung tendenziell stärker zustimmten als in der off-Bedingung, während es sich bei den Novizen genau umgekehrt verhielt.

6.3.3.2 Natürlichkeitsbeurteilung

Zur Beurteilung der Natürlichkeit des Systems wurden die KU-Einzelitems zur Erfassung der Sympathie, der Höflichkeit, der Intuitivität und der Menschenähnlichkeit des Systems herangezogen und in eine multivariate Varianzanalyse mit Messwiederholung einbezogen.

Die Dialogbedingungen zeigten auf das Konstrukt der Natürlichkeit einen tendenziell signifikanten Einfluss (F[8,68]= 2.2; p= .04; η^2= .21), der sich nicht in univariaten Effekten oder in dem Natürlichkeitskoeffizient abbilden ließ. So wurden alle vier Eigenschaftsdimensionen über die Dialogbedingungen hinweg als *gut* eingestuft.

Betrachtete man den Faktor **Erfahrung**, so zeigte sich ein Haupteffekt des Zwischensubjektfaktors auf das Konstrukt Natürlichkeit (F[4,15]= 3.4; p= .04; η^2= .47). Dabei wurde jede Facette der Natürlichkeit unabhängig von der Dialogbedingung von den Experten besser bewertet. Hinsichtlich der Sympathie konnte ein Interaktionseffekt zwischen den Dialogbedingungen und der Erfahrung beobachtet werden (F[1.5,26.5]= 6.6; p= .01; η^2=.27). Während Novizen das wrong-System sympathischer als das angleichende System einstuften, verhielt es sich bei den Experten genau umgekehrt (siehe Abbildung 51).

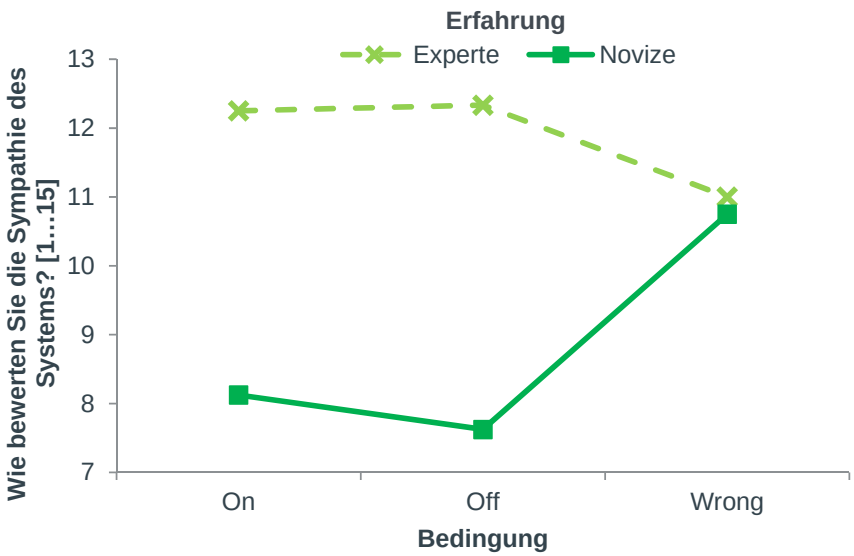

Abbildung 51: Sympathiebeurteilung (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Dialogbedingung und der Erfahrung

6.3.3.3 Konsistenzbeurteilungen

Die Konsistenzbeurteilung zeigte keinen signifikanter Effekt des Innersubjektfaktors Dialogbedingung (F[1.5,27.9]= 2.4; p= .12). Bei heuristischer Betrachtung der Innersubjektkontraste fand sich ein Effekt zwischen der Konsistenzbeurteilung in der off- und on-Bedingung (F[1,19]= 7.6; p= .01; η^2= .28). So wurde die Konsistenz in der Dialogbedingung mit Alignment signifikant besser eingeschätzt als in der statischen Dialogbedingung.

6.3.3.4 Sprachkompetenzbeurteilung

Zur Analyse der nutzerseitig eingeschätzten sprachlichen Kompetenz des Systems wurden die KU-Einzelitems zur Erfassung der Wortschatzgröße, der Wortgewandtheit, der Grammatikkenntnisse und der Fähigkeit des Systems Sprache zu verstehen, herangezogen. Die vier Facetten der Sprachkompetenz wurden in eine multivariate Varianzanalyse mit Messwiederholung einbezogen, um den Einfluss der Dialogbedingungen zu berechnen.

Es ließ sich kein Effekt der Dialogbedingungen auf das Konstrukt Sprachkompetenz beobachten (F[8,68]< 1; p= .52). Obwohl die Wortschatzgröße in der on-Bedingung angemessener als in der off- oder wrong-Bedingung beurteilt wurde, haben sich diese Unterschiede als nicht statistisch signifikant erwiesen (F[2,38]< 1; p= .54).

Betrachtete man die **Erfahrung** der Probanden, so zeigte sich zunächst ein Haupteffekt des Zwischensubjektfaktors auf alle Dimensionen der sprachlichen Kompetenz (siehe Tabelle 37). Insgesamt schätzten Probanden mit Erfahrung die Systemkompetenz über alle Dialogbedingung besser ein als Probanden ohne Erfahrung.

Tabelle 37: Effekte Erfahrung auf sprachliche Kompetenz

	Effekt Erfahrung		
	$F_{(1,18)}$	p	η^2
Wortschatz	**8.0**	**.01**	**.31**
Wortgewandtheit	**9.7**	**.01**	**.35**
Grammatikverständnis	**9.1**	**.01**	**.34**
Sprachverstehen	**6.5**	**.02**	**.27**

Darüber hinaus zeigte sich ein Trend bei dem Interaktionseffekt der Dialogbedingung mit der Erfahrung (F[8,64]= 1.8 p= .09; η^2= .19). Bei Betrachtung der Innersubjektkontraste zeigte sich, dass Probanden mit Erfahrung die Grammatikkenntnisse des wrong-Systems schlechter ein-

schätzten als die des on-Systems. Für Probanden ohne SDS-Erfahrung verhielt sich dies genau umgekehrt.

Während die Novizen auch die Fähigkeit zum Sprachverstehen für off-System höher einschätzten als für das on-System, verhielt sich dies bei den Experten genau gegenläufig.

6.3.3.5 Ablenkung

Zur Analyse der Ablenkung von der Fahraufgabe wurden in der vorliegenden Untersuchung vier Parameter betrachtet. Neben der Spurablage, der Standardabweichung des Abstands und der Geschwindigkeit wurde auch die subjektive Einschätzung der Ablenkung analysiert. Im Rahmen der Fahrdaten wurde sich auf die um die Baseline bereinigten Werte bezogen.

Die Dialogbedingungen zeigten keinen signifikanten Einfluss auf das Konstrukt Ablenkung (F[8,52]< 1; p= .96).

Insgesamt konnte nicht für jede Nutzergruppe belegt werden, dass der Transfer von Angleichungsprozessen auf linguistischer Ebene dazu beitragen kann, die Nutzerfreundlichkeit eines Sprachdialogsystems zu erhöhen. Im Folgenden werden die Ergebnisse diskutiert.

6.3.4 Diskussion

Angelehnt an den Interactive Alignment Ansatz (Pickering & Garrod, 2004) wurde ein System implementiert, das sich sowohl lexikalisch als auch teilweise syntaktisch an die Nutzereingabe anpasste. Darauf aufbauend wurde eine Untersuchung realisiert, die es ermöglicht, die positiven Effekte der Angleichung auch abseits von Effizienzsteigerung oder Tutoriumsszenarien im Fahrtkontext zu betrachten.

6.3.4.1 Diskussion der Ergebnisse

Der Manipulations-Check konnte belegen, dass nahezu alle Nutzereingaben in der on-Bedingung durch das System angeglichen werden konnten. Auch wenn aufgrund dieser Befundlage von einer erfolgreichen Manipulation des Dialogverhaltens ausgegangen werden kann, waren die Unterschiede den Versuchspersonen häufig nicht bewusst. So konnten auch nach einer expliziten Instruktion, dass sich die Systeme hinsichtlich ihrer Wortwahl unterscheiden, lediglich vier Probanden die Unterschiede zwischen den Dialogbedingungen korrekt benennen. Bei Anwendung einer strengeren Definition zeigte nur ein Proband **Alignment Awareness**. Auch wenn das Bewusstsein der Angleichung damit als gering einzustufen ist, so kann doch eine gewisse Sensitivität der Versuchspersonen gegenüber der Manipulation bei Betrachtung aller Systembeurteilungen beobachtet werden. Viele Probanden erkannten Auswirkungen der Dialoggestaltung auch in

der off- und wrong-Bedingung und nannten beispielsweise zu monotone oder verwirrende Rückmeldungen als Flops der Systeme. Es kann in gewisser Weise davon ausgegangen werden, dass die Unterschiede in der Wortwahl wahrgenommen wurden, auch wenn sie nicht immer als Bewertungsgrundlage verbalisiert werden konnten. Hinsichtlich der **Präferenzmessung** konnte belegt werden, dass insbesondere Probanden sich häufiger für das on-System entschieden, die sich der Angleichung des Systems bewusst waren.

Zur Beantwortung der Forschungsfragen war es weiterhin von Interesse, dass sich die Dialogbedingungen nicht hinsichtlich ihrer **Effizienz** unterschieden. So konnten keine Differenzen bei den objektiven und subjektiven Effizienzparametern für die verschiedenen Dialogbedingungen festgestellt werden. Eine bewusste Unterscheidung zu der Untersuchung von Porzel (2006) konnte dadurch realisiert werden.

Entgegen der Annahmen konnte jedoch keine Zunahme der **Gebrauchstauglichkeitsbeurteilung** durch das Alignment beobachtet werden. Das Konstrukt der Usability zeigte keine Beeinflussung durch das Dialogverhalten. Es konnte lediglich ein Trend festgestellt werden, dass das konstant reagierende System am schlechtesten bewertet wurde. Dennoch erzielte das wrong-System ähnlich hohe Usability-Werte wie das on-System.

Dieses Bewertungsschema manifestierte sich bei der subjektiven Beurteilung der **Nutzungsbereitschaft**. So zeigten sich bei der Bewertung der Fehleranfälligkeit des Systems nur in der on- und wrong-Bedingung positive Effekte gegenüber der Prä-Messung. Das konstant reagierende System führte zu keiner signifikanten Reduktion der Fehleranfälligkeitsbeurteilungen verglichen zur Prä-Interaktionsmessung. Durch den großen Wortschatz, der dem on- und dem wrong-System gemein war, gingen Nutzer davon aus, dass diese beiden Systeme auch mehr Worte verstehen könnten als ein System, welches immer nur eine gleichbleibende Formulierung verwendete. In verschiedenen Untersuchungen konnte bereits nachgewiesen werden, dass die Wortwahl der Nutzer stark von den Systemausgaben abhängt (vgl. Brennan, 1991; Zoltan-Ford, 1991). An dieser Stelle zeigten sich darauf aufbauend Hinweise, dass die Sprachausgaben auch zur Einschätzung der Systemkompetenz herangezogen wird. Das Promptdesign übt damit aktiv Einfluss auf das Rezipientenmodell aus, welches Nutzer von ihrem Gesprächspartner ausbilden. Abgesehen von den positiven Effekten des Alignments auf die Beurteilung der Fehleranfälligkeit, konnte die Annahme der erhöhten allgemeinen Nutzungsbereitschaft durch die Angleichung des Systems somit nicht bestätigt werden.

Ebenso zeigten die **Zufriedenheitsbeurteilung** und die subjektive Präferenz der Probanden für ein Bediensystem ebenfalls ein nicht mit der Hypothese übereinstimmendes Ergebnis. Die Probanden entschieden sich in einem beinah gleichen Ausmaß für die unterschiedlichen Dialogbedingungen. Ihre Zufriedenheitsratings gestalteten sich unabhängig von der Dialogbedingung.

Insgesamt musste auch für alle objektiven Parameter der **Effektivität** die Ursprungshypothese verworfen werden. Weder die Anzahl der Fehlerkennungen, noch der Bedienfehler reduzierten sich durch die lexikalische Angleichung des Systems. Lediglich im Rahmen der subjektiven Effektivitätsbeurteilung konnte die Annahme bestätigt werden; dass das adaptive System tendenziell **zuverlässiger** eingestuft wird als das unangepasst reagierende wrong-System. Diese Tendenz steht in Einklang zu dem bereits beschriebenen Effekt der Fehleranfälligkeit im Rahmen der Nutzungsbereitschaft und der gesamten Collaborative Theory (Clark, 1996). So erscheint Nutzern die Kommunikation durch das Alignment zielgerichteter, da der geteilte Common Ground innerhalb der on-Bedingung deutlicher betont und bestätigt wird. Bei einer systemseitigen Angleichung kann sich der Nutzer des gegenseitigen Verständnisses sicherer sein als bei Systemen, die bewusst andere Formulierungen verwenden. Wenn das System keine Angleichung zeigt und andere Formulierungen verwendet, ist ein Missverständnis wahrscheinlicher und das System wirkt unzuverlässiger.

Weiterhin konnte keine Vermeidung von unnötigen **Korrekturen** durch das angleichende System beobachtet werden. Die Kommandos waren semantisch alle im selben Bereich verortet und enthielten kaum Mehrdeutigkeiten. Ein Missverständnis, wie in dem Verkehrsfunkbeispiel von Hassel (2006), konnte nicht beobachtet bzw. verhindert werden. Insgesamt stellt jenes Beispiel einen sehr selten zu beobachtenden Einzelfall dar, der sich experimentell nur schwer herstellen lassen kann.

Auch im Bereich der Fehlbedienungen konnte keine Reduktion der **TTE-Fehler** durch die Angleichung beobachtet werden. Nenkova (2008) fand zwar korrelative Zusammenhänge, dass in zwischenmenschlichen Dialogen Unterbrechungen bei lexikalischer Angleichung weniger häufig auftraten, diese lassen sich jedoch nur bedingt kausal interpretieren. So sollte in Betracht gezogen werden, dass es in zwischenmenschlichen Gesprächen immer dann zu einer Unterbrechung kommt (Nenkova trennt zwischen Sprachüberlappungen und Unterbrechungen), wenn ein Missverständnis vorliegt. Dies ist bei einer lexikalischen Angleichung äußerst seltener der Fall. Somit treten Unterbrechungen und Angleichungen zwar selten gemeinsam auf, aber ein kausaler Zusammenhang besteht nicht. So führen die Angleichungen nicht zu weniger Unterbrechungen, sondern das Alignment deutet eher auf ein korrektes Verständnis hin, das keiner Unterbrechung bedarf.

Darüber hinaus konnte Nenkova (ebd.) auch korrelative Zusammenhänge zwischen linguistischem Alignment und Sprachüberlappungen feststellen. So stehen diese Überlappungen charakteristisch für gute Koordination und eine lebendige Konversation und scheinen häufig zusammen mit lexikalischer Angleichung stattzufinden. Es kann davon ausgegangen werden, dass Sprecher, wenn sie sich dem gegenseitigen Verständnis durch das Entrainment sicher sind, schneller im Dialog fortfahren. In der vorliegenden Untersuchung konnte zwar nicht bestätigt werden, dass zu

frühe Spracheingaben durch die Angleichung provoziert wurden, es konnte allerdings beobachtet werden, dass Nutzer teilweise unzufrieden mit der **Dialoglänge** waren und Unmut darüber äußerten, der Wiederholung ihrer eigenen Formulierung zuzuhören. Ohne die Möglichkeit, das System an dieser Stelle zu unterbrechen und damit von der Angleichung im Hinblick auf die Effizienz zu profitieren, scheint das Alignment für die Nutzer eher kontraproduktiv. In späteren Untersuchungen sollte den Nutzern explizit die Möglichkeit eingeräumt werden, ein System sprachlich zu überlappen und somit bei etablierten CG effizienter fortzufahren. Nur so können die Vorteile der Angleichung zum Tragen kommen.

Weiterhin zeigten sich alle Aspekte der **Natürlichkeit** unabhängig von den Dialogbedingungen. Damit konnten die erwarteten positiven Effekte des Alignments auf die Dimensionen Sympathie, Höflichkeit, Menschähnlichkeit und Intuitivität nicht beobachtet werden. Auch wenn Nenkova (2008) belegte, dass eine lexikalische Angleichung ein guter Indikator für wahrgenommene Natürlichkeit der Unterhaltung ist, so lässt sich auch dies nicht kausal interpretieren.

Erwartungskonform konnte bestätigt werden, dass ein angleichendes System zu keinen negativen Effekten bei der **Konsistenzbeurteilung** führt. So zeigte sich sogar, dass das adaptive System als konsistenter bewertet wurde, als das immer gleich reagierende System. In Bezug auf das in Studie II bereits diskutierte Konsistenzdilemma bedeutet dies, dass ein natürlich-reagierendes adaptives System hier sogar als konsistenter wahrgenommen wurde, als ein statisch reagierendes System. Konsistenz im Rahmen der Gestaltung von SDS sollte demnach nicht bedeuten immer gleich zu reagieren, sondern nutzer- und situationsangemessene Anpassungen zuzulassen.

Entgegen der Annahmen konnte kein positiver Effekt des Alignments auf das Konstrukt **Sprachkompetenz** beobachtet werden.

Für die Analyse der subjektiven **Ablenkungsbeurteilung** und der Fahrdaten zeigten sich ebenfalls keine mit der Annahme übereinstimmenden Ergebnisse. So konnten weder für die objektiven Fahrparameter, noch für die subjektive Ablenkungsbeurteilung signifikante Differenzen festgestellt werden. Obwohl sich bei der Analyse der Spurhaltedaten in der wrong-Bedingung die größte, über alle Aufgaben gemittelte Standardabweichung der Querposition zeigte, wurde dieser Unterschied nicht signifikant. Auch die Beurteilung der Anstrengungsdimension innerhalb des ITU zeigte sich innerhalb der Aufgabenfahrten als unabhängig von dem Dialogverhalten. Es konnte nicht belegt werden, dass sich der kognitive Aufwand der Sprachbedienung durch das Alignment reduziert und die Dialoge mehr automatische Prozesse bedienen.

Insgesamt konnten die erwarteten positiven Effekte des Alignments nicht nachgewiesen werden. Eine mögliche Erklärung dafür kann in einem Deckeneffekt des Verständnisses liegen. Denn be-

trachtet man die Synchronisation der Situationsmodelle, so ist nach Garrod und Clark (1993) eine Dialogdyade immer dann ausgeglichen, wenn Sprecher A etwas äußert, das konsistent mit Empfänger B‘s aktivierten, semantischen und pragmatischen Repräsentationen des Dialogs ist und umgekehrt. Im Grunde hat dies in allen Dialogbedingungen stattgefunden. Zumindest das semantische Situationsmodell nach Pickering und Garrod (2004) wurde in jeder Dialogbedingung erfolgreich angeglichen, auch wenn dies nicht immer mit einer Synchronisation auf lexikalischer Ebene einherging. Somit kann ein Deckeneffekt des gegenseitigen pragmatischen Verständnisses vorliegen, der durch eine zusätzliche Angleichung nicht erhöht werden kann.

Neben den Effekten auf die Zuverlässigkeits- und Kompetenzbeurteilung zeigte ein Nichtangleichen auch keine weiteren negativen Konsequenzen. Anders als in zwischenmenschlichen Dialogdyaden scheinen Nutzer nicht mit der Erwartungshaltung in einen Dialog mit einem SDS zu kommen, dass bereits akzeptierte Begriffe verwendet werden (Metzinger & Brennan, 2003). Trotz des fehlenden Alignments wirkte das System weder unsympathischer, inkompetenter oder unhöflicher. Es ließen sich auch keine Effekte der erhöhten nutzerseitigen Beanspruchung erleben. Anders als in zwischenmenschlichen Dialogen (Branigan et al., 2010) führt die Wahl von alternativen Äußerungen demnach nicht zu **negativen Affekten** des Gesprächspartners. Als Erklärung hierfür ist einerseits denkbar, dass die Effektivität und Effizienz eines Mensch-Maschine-Dialogs sehr viel bedeutsamer ist, als der affektive Beziehungsaufbau zu einem sprachverarbeitenden System. Nutzer, die in Interaktionen mit Sprachdialogsystemen selbst unhöflichere Verhaltensweisen zeigen als in zwischenmenschlichen Gesprächen, reagieren verzeihlicher, wenn ein SDS nicht die Etikette wahrt.

Exkurs: Innerhalb der **Fahrdatenanalysen** konnte beobachtet werden, dass die Versuchspersonen zum Teil bessere Spurhaltefähigkeiten während der Aufgabenfahrten, verglichen zu der Referenzfahrt, zeigten. Ähnliche Effekte berichtete bereits Kaiser (2009), die 15 Studien zusammenfassend dahingehend untersuchte, wie kognitive im Vergleich zu visuell-manuell beanspruchenden Nebenaufgaben die Querführung beeinträchtigen. Während die visuell-manuellen Nebenaufgaben die Standardabweichung der lateralen Position (SDLP) erhöhten, führten kognitive Zweitaufgaben eher zu geringeren Spurabweichungen. Die Verbesserung der Querführung während der Durchführung kognitiver Aufgaben, verglichen zur Kontrollfahrt, begründet sie durch die Theorie der reaktiven Anspannungssteigerung von Düker (1963). Im Rahmen dieser Theorie führt die zusätzliche Durchführung einer kognitiven Aufgabe während der Fahrt zu einer Erhöhung der psychischen Anspannung, die sich positiv auf die Querführung auswirkt. Bei der Interaktion mit einem SDS kann der Fahrer Veränderungen der Position seines Fahrzeugs auf der Straße sehen, ist nicht motorisch eingeschränkt und kann sich optimal auf die Spurhaltung konzentrieren. Die Fahraufgabe profitiert demnach durch die erhöhte psychische Anspannung durch die Zweitaufgabe. Caird (2008) fand im Rahmen einer meta-analytischen Betrachtung von Handynutzung während der Fahrt widersprüchliche Effekte bei dem Vergleich zwischen der Ba-

seline und den Aufgabenfahrten. Kritisch zu betrachten ist an dieser Stelle der Fakt, dass eine hohe Anspannung bei Hinzunahme einer weiteren Anforderung schnell zu einer Überlastung des Fahrers führen kann.

Bezog man die **Erfahrungswerte** der Probanden mit Sprachdialogsystemen mit in die Analysen ein, so können einige der Annahmen für die Expertenstichprobe bestätigt werden.

Zunächst zeigten sich auf vielen Variablen Haupteffekte des Erfahrungslevels. Insgesamt beurteilten Probanden mit Erfahrung die Dialogsysteme über alle Bedingungen hinweg zuverlässiger und natürlicher. Alle Dimensionen der sprachlichen Kompetenz und nahezu alle Facetten der Nutzungsbereitschaft wurden von erfahrenen Nutzern besser bewertet. Insgesamt waren die Nutzer mit Expertise auch zufriedener mit den drei Umsetzungen. Diese Befundlage verdeutlicht eindrucksvoll die Lücke zwischen der nutzerseitigen Erwartung und den technischen Möglichkeiten. Da Experten bereits Erfahrungen mit derzeitigen automotiven SDS sammeln konnten, beurteilten sie die Forschungsprototypen auf nahezu jeder Dimension besser als die Novizen. Letztere stellten aufgrund von zu wenig alltäglicher Erfahrung mit sprachverarbeitenden Systemen zu hohe Ansprüche an die Technik.

Neben den Haupteffekten zeigten sich auch interessante Interaktionen des Erfahrungslevels mit den Dialogbedingungen. Während beispielsweise Novizen mit der on-Bedingung **unzufriedener** waren als mit der wrong-Bedingung, verhielt es sich bei den Experten exakt umgekehrt. Dieses Bewertungsschema fand sich auch bei der Sympathiebeurteilung. Während Novizen das System, das bewusst andere Formulierungen verwendete, **sympathischer** als das angleichende System einstuften, verhielt es sich bei den Experten genau umgekehrt. Innerhalb der Expertenstichprobe konnte demnach die Annahme, dass ein adaptives System sympathischer ist und die Nutzerzufriedenheit steigt, bestätigt werden. Somit zeigten nur die Probanden mit Erfahrung bei der Nutzung von SDS affektive Reaktionen, wie sie in den Untersuchungen von zwischenmenschlichen Dialogen berichtet werden.

Weiterhin fanden sich im Rahmen der Nutzungsbereitschaft Effekte, die für eine teilweise Bestätigung der Hypothese in der Expertenstichprobe sprachen. So zeigt sich, dass Probanden mit Erfahrung die **Grammatikkenntnisse** des wrong-Systems schlechter einschätzten, als die des on-Systems. Für Probanden ohne SDS-Erfahrung verhielt sich dies genau umgekehrt. Während die Novizen auch die Fähigkeit zum **Sprachverstehen** für das off-System höher einschätzten als für das on-System, verhielt sich dies bei den Experten genau gegenläufig. Dies stützt die Annahme, dass erfahrene Nutzer aus dem Wortschatz des Systems dessen Kompetenz zum Sprachverstehen ableiten.

Diese Befundlage kann dafür sprechen, dass Experten und Novizen bei der Beurteilung von Dialogen unterschiedliche **Bewertungsheuristiken** zugrunde legen. Während Novizen einen größeren Wortschatz immer positiv bewerteten, bemerkten Experten feinfühliger Differenzen. Gegebenenfalls kann man ihnen die Fähigkeit zutrauen, den erlebten Dialog auf die reale Interaktion während der Fahrt zu antizipieren und damit besser einschätzen zu können, wie ablenkend und mental belastend einzelne Formulierungen des wrong-Systems sein könnten.

Diese Interpretation kann durch Antworten der Versuchspersonen auf die Frage nach den Tops und Flops der Systeme gestützt werden. Während sich die Experten bei der Einschätzung des adaptiven Systems auf die Höflichkeit und die Wortgewandtheit bezogen, nannten die Novizen vorrangig die Prosodie und die gute Erkennrate als Beurteilungsgrundlage. Auch bei der Beurteilung des wrong-Systems fanden sich unterstützende Antworten. So nannten die Experten als Flops des Systems, dass das System weniger höflich sei, störende Widerworte gebe und die abweichende Wiederholung der Eingaben ablenkend sei. Dagegen fokussierten die Novizen auf die Erkennraten und Bestätigungsanfragen des Systems, die nicht Gegenstand der Manipulation waren und über alle Bedingungen gleich vorlagen (vgl. Kap. 6.3.3.1).

Während in Porzels (2006) Studie insbesondere die Novizen von der Angleichung des Systems profitierten, waren es in der vorliegenden Untersuchung eher die Experten. Da Porzels System einen tutoriellen Charakter hatte, konnten Novizen mit einem angleichenden Dialog ein schnelleres Verständnis und damit das Dialogziel erreichen. Dagegen ging es in der vorliegenden Studie primär um einen strategischen Gestaltungsaspekt, dessen Nutzen Probanden mit SDS-Erfahrung besser beurteilen konnten.

In der vorliegenden Untersuchung konnten die positiven Effekte linguistischen Alignments bei der Mensch-Maschine-Interaktion nicht in dem erwarteten Ausmaß gezeigt werden. Als Begründung für die fehlenden oder geringen Effekte der lexikalischen Angleichungen sollen an folgender Stelle verschiedene Umsetzungsdetails diskutiert werden.

6.3.4.2 Diskussion Umsetzung

In der Ausgangsstudie von Porzel et al. (2006) erzielte das adaptive System allseits bessere Beurteilungen als das konstant reagierende System. So fielen die Bewertung des Wortschatzes und der Gebrauchstauglichkeit und die generelle Akzeptanz für das adaptive System besser aus. Diese Befunde konnten in der vorliegenden Studie nicht repliziert werden.

Innerhalb des Problemlösedialogs der WoZ-Studie (ebd.) konnte eine Angleichung an den Nutzer durch die Expertenrolle des Sprachdialogagenten ein schnelleres beidseitiges Verständnis herbeiführen. Die lexikalische Angleichung konnte darüber hinaus als implizite Bestätigung der Dia-

logschritte genutzt werden, um so den Dialogablauf signifikant zu beschleunigen. Während also in der Ausgangsstudie (ebd.) die Zeit bis zur Aufgabenerfüllung durch das lexikalische Alignment um durchschnittlich eine Minute reduziert werden konnte, erfolgte in der vorliegenden Studie per Manipulationsdefinition keine Steigerung der **Effizienz** durch die Angleichung. Es handelte sich nicht um eine Art Tutoriumszenario, sprich dem SDS kam keine Expertenrolle bei der Ausführung der Kommandos zu und die expliziten Bestätigungsanfragen wurden unabhängig von der Angleichung präsentiert.

Es lässt sich ableiten, dass bei einer Konstanthaltung des Faktors Effizienz die positiven Effekte des Alignments auf die Gebrauchstauglichkeit und Akzeptanz nicht repliziert werden können. Das on-System wurde in der Gesamtstichprobe weder sympathischer noch gebrauchstauglicher eingeschätzt. Es liegt die Interpretation nahe, dass die strategischen Vorteile des Alignments in Mensch-Maschine-Interaktionen nur in Kombination mit Effizienzverbesserungen auftreten. Somit werden angleichende Systeme nur dann positiver beurteilt, wenn sie die Zielerreichung beschleunigen oder begünstigen können. Ohne eine Effizienz oder Effektivitätssteigerung durch das Alignment, finden sich auch keine positiven Effekte auf die nutzerseitige Systembewertung. Dies kann einerseits auf die Relevanz der Stellgröße Effizienz im Dialog zurückgeführt werden (vgl. Studie II). Andererseits kann diskutiert werden, ob der strategisch-sozialen Komponente im Mensch-Maschine-Dialog im Fahrzeug eine geringere Bedeutung zukommt. So werden Sprachdialogsysteme nur bedingt als Gesprächspartner verstanden, die durch bestimmte Verhaltensweisen sympathischer erscheinen können.

Es muss auch diskutiert werden, ob die Untersuchung des Alignments ohne Effizienzsteigerung zielführend ist. So konnte Kapitel 3.2.3 zeigen, dass Angleichungsprozesse zwischenmenschlich als Verstehensevidenz fungieren und stets zu einer schnelleren Etablierung und Bestätigung des CG beitragen. Zwischenmenschlich kann linguistisches Alignment ohne Zeiteinsparung kaum beobachtet werden. Eine losgelöste Betrachtung des Alignments von der Effizienzsteigerung diente hier der Untersuchung der strategischen Funktion der Angleichung. Für die Implementation in Seriensysteme sollte allerdings der Effizienzaspekt nicht außen vor gelassen werden.

Weiterhin gilt zu diskutieren, ob der Ebene der **phonetischen Angleichung** zu wenig Beachtung geschenkt wurde. Insbesondere in der on-Bedingung konnte sehr viel Kritik an der Qualität der Sprachausgabe verzeichnet werden. Insgesamt sechs Probanden kritisieren die unnatürliche Aussprache des Systems. Da die TTS in allen drei Bedingungen identisch waren, war es überraschend, dass diese Kritik in den beiden anderen Bedingungen kaum geäußert wurde. In der wrong-Bedingung, in der die gleichen Sprachausgaben wie in der on-Bedingung zu anderen Zeitpunkten präsentiert wurden, konnten keine vergleichbaren Rückmeldungen dokumentiert werden. Eine *abgehackte und unnatürliche Stimme* wurde lediglich von einer Person innerhalb der wrong-Bedingung erwähnt. Daher kann angenommen werden, dass das Fehlen einer linguistischen

Angleichungsebene durch die VP in der vorliegenden Untersuchung betont wahrgenommen wurde, da die Dimensionen des Alignments in zwischenmenschlichen Dialogen immer gemeinsam auftreten. Indem innerhalb der on-Bedingung auf der syntaktischen und lexikalischen Ebene eine Angleichung stattfand, wirkte sich die fehlende Angleichung auf phonetischer Ebene besonders negativ aus.

Es sollte tiefergehend betrachtet werden, ob eine erfolgreiche Angleichung auch bei Dialogsystemen alle linguistischen Ebenen bedienen, um die Priming-basierten Vorteile zu nutzen. Eventuell ist gerade die phonologische Angleichung zur Kostenreduktion im Bereich der Rezeption eine zentrale Stellgröße. Somit wäre auch eine ergänzende Studie denkbar, die eine vollständige systemseitige Angleichung auf allen Ebenen des Interactive Alignment Ansatzes (Pickering & Garrod, 2004) adressiert. Die vorliegenden Befunde geben Grund zur Annahme, dass die fehlende Angleichung auf phonetisch-phonologischer Ebene störend wirkte. Daher sollte geprüft werden, ob der positive Einfluss der unvermittelten Priming-Prozesse auf die Fahreraufmerksamkeit bei einer Angleichung auf allen linguistischen Ebenen stärker beobachtet werden kann. Weiterhin sollten fortführende Untersuchungen angeregt werden, die die Relevanz der einzelnen Ebenen der linguistischen Angleichung definieren. So deutet die Ergebnislage darauf hin, dass die einzelnen Ebenen nicht nur ergänzende Wirkung besitzen, wie von Pickering und Garrod (2004) postuliert, sondern nur dann vollständig sind, wenn Alignment auf allen linguistischen Ebenen stattfindet.

Weiterhin zeichnete sich im Rahmen der Fahrdatenanalyse bei der SDLP ein Bodeneffekt aller Dialogbedingungen ab. Die simulierte Aufgabenfahrt wurde weniger von der sprachlich-kognitiven Aufgabe beeinflusst, als dies im realen Straßenverkehr der Fall wäre (Ranney, 2008). Die Befundlage von Kaiser (2009) lässt vermuten, dass die kognitive Beanspruchung, die durch die Interaktion mit dem Sprachdialogsystem entsteht, bei der alleinigen Betrachtung der Spurhaltung unterschätzt wird. Dies könnte ein Hinweis darauf sein, dass die etablierten Fahrparameter in einem simulierten Folgefahrtszenario nicht trennscharf genug zwischen Dialoggestaltungsalternativen unterscheiden konnten. Zur besseren Erfassung der kognitiven Beanspruchungseffekte des SDS auf die Fahreraufmerksamkeit sollten in Zukunft Reaktionstests wie der peripheral detection task[34] in Ergänzung zu den üblichen Fahrparametern in Betracht gezogen werden.

[34] Bei der peripheral detection task werden während der Fahrt alle drei bis sechs Sekunden für maximal zwei Sekunden visuelle Reize dargeboten, auf die durch Tastendruck reagiert werden soll (Martens & van Winsum, 2000). Sie stellt ein gutes Maß der kognitiven Beanspruchung dar.

Auch hier wirkt sich, wie bereits in Studie I und II erläutert, die **Künstlichkeit der Laborsituation** in mehreren Aspekten auf die Mensch-SDS-Interkation aus. Durch das Vorsprechen der Kommandos wurde ein großer Teil der Konzeptualisierungs- oder Formulierungsphase (Kap. 3.1.2.2) der Sprachproduktion vorweggenommen. Faktisch übernahm der VL die Ausformulierung der Botschaft für den Probanden. Dieser sprach das Kommando nur nach. Im Grunde bestand somit ein Dialog zwischen drei Instanzen, bei denen bereits eine Angleichung zwischen dem Versuchsleiter und dem Probanden erfolgte und das Sprachdialogsystem nur die dritte Instanz bildete. Auch wenn dies eine hohe Standardisierung bei der Durchführung der Untersuchung ermöglichte, ließen sich unter Umständen stärkere Effekte des systemseitigen Alignments nachweisen, wenn die Ausformulierung des Kommandos bereits beim Probanden läge und dieser damit vollständig die Kosten für die Konzeptualisierung und Formulierung tragen würde. Damit würden die Versuchspersonen gegebenenfalls mehr von der Angleichung bei der Rezeption profitieren. Zukünftige Untersuchungen sollten mehr Wert auf die eigenständige Formulierung der Kommandos durch die Probanden legen.

Durch das Potential zur Zeiteinsparung in Problemlösedialogen lassen sich auch im Rahmen von automotiven Sprachdialogsystemen **Bereiche** vermuten, in denen linguistisches Alignment größeres Potential als in dem vorliegenden Untersuchungskontext aufweist. Bereits Core und Moore (2004) wiesen darauf hin, dass die lexikalische Angleichung insbesondere bei Tutor-Systemen von großer Bedeutung ist. Auch die vieldiskutierten Erkenntnisse von Porzel et al. (2006) lassen darauf schließen, dass es im Rahmen von Problemlösedialogen lohnenswert ist, den technischen Aufwand der lexikalischen Adaption zu betreiben. Während die positiven Effekte der Angleichung in dem Telefonkontext nur gering waren, wäre es vorstellbar, dass sie in Hilfekontexten oder sprachbedienbaren Bordbüchern größeres Potential aufweisen würden.

6.3.4.3 Diskussion Methode

Zur Diskussion der Vor- und Nachteile des **Messwiederholungsdesigns** sei auf die Studie II verwiesen (Kap. 6.2.4.3). Zusätzlich sei erwähnt, dass das reine Messwiederholungsdesign in der vorliegenden Untersuchung dazu führte, dass Personen aufgrund einer fehlenden Bedingung aus der gesamten Analyse ausgeschlossen wurden. Traten also technische Probleme bei der Erhebung der letzten Bedingung auf, so konnten auch die Daten der bereits aufgezeichneten vorangegangenen Bedingungen nicht Verwendung finden.

Weiterhin ließen sich Reihenfolgeeffekte nicht vollständig ausschließen. Durch ein vollständiges Ausbalancieren wurde versucht, Ermüdungs-, Übungs- und Carry-over-Effekte über die Versuchsbedingungen auszugleichen. Aufgrund der Kombinationsmöglichkeiten ergaben sich sechs Reihenfolgegruppen, die aufgrund ihrer geringen Gruppenbesetzung Tests auf Reihenfolgeeffekte nicht ermöglichten. Für weitere Studien sollte daher in Erwägung gezogen werden, die Befun-

de in komplett randomisierter Vier-Felder-Form und mit einer a posteriori determinierter Stichprobengröße zu überprüfen.

Weiterhin muss methodisch angemerkt werden, dass Effektstärken mit der **Manipulation** der unabhängigen Variablen zusammenhängen. Je größer die Manipulation ausfällt, desto stärkere Effekte sind zu erwarten (Brandstätter, 1999). Da die Manipulation des Dialogverhaltens sich hier nur auf die Wortwahl bezog, kann sie als gering eingestuft werden. Ein Deckeneffekt der hohen Qualität der drei Systeme lässt sich insgesamt nicht ausschließen.

Neben der zuvor diskutierten Ergänzung der Fahraufgabe sollte auch die Erhebung der subjektiven **Ablenkungsbeurteilungen**, die bisher lediglich durch Einzelitems erfolgte, ausgebaut werden. Durch die Erhebung eines Fragebogens zu den Aspekten der Beanspruchung (z. B. DALI) könnten Bewertungen über unterschiedlich gepolte Items differenzierter erfasst und eine reliablere und validere Erhebung der Beanspruchung durch die Sprachdialoge ermöglicht werden.

Die fehlenden Effekte könnten teilweise auch mit dem geringen **Erfahrungslevel** der Probanden mit natürlich-sprachlichen Bedienungsapplikationen und deren unterschiedlichen Bewertungsheuristiken erklärt werden. Die Experten-Novizen Re-Analyse gibt einen ersten Hinweis darauf, dass Probanden mit Erfahrung ein System, welches linguistisches Alignment zeigt, besser beurteilen. Es kann davon ausgegangen werden, dass sich aufgrund der kleinen Expertenstichprobe und der daraus folgenden relativ geringen Teststärke weitere bedeutsame Effekte nicht signifikant zeigten (Bortz, 2005). Umso mehr kann aber auch die Relevanz der diskutierten Einflüsse betont werden. Es sei demnach angeraten, in fortführenden Untersuchungen die Fragestellung erneut mit einer ausreichend großen Expertenstichprobe zu beleuchten.

In zukünftigen Untersuchungen sollte weiterhin überprüft werden, ob sich konsistentere Befunde finden lassen, wenn direkt nach jeder Systeminteraktion nach der mentalen Belastung gefragt wird. Auch das Einbeziehen einer Videoanalyse scheint ein vielversprechender Untersuchungsansatz.

6.3.5 Zusammenfassung

Es konnte gezeigt werden, dass die Umsetzung linguistischen Alignments in Anlehnung an das Interactive Alignment Modell (Pickering & Garrod, 2004) nur bedingt eine Erleichterung des Grounding-Prozesses mit einem Sprachdialogsystem ermöglichte. Der bestehende Forschungsstand konnte um Erkenntnisse ergänzt werden, wie sich die lexikalische und syntaktische Angleichung eines automotiven Sprachdialogsystems auf die nutzerseitige Systembewertung auswirken.

Ohne eine gleichzeitige Verringerung des zeitlichen Aufwandes durch das Alignment konnte keine Steigerung der allgemeinen Gebrauchstauglichkeit nachgewiesen werden. Die strategischen Vorteile des Alignments konnten nur innerhalb der Expertenstichprobe beobachtet werden. Dennoch kann durch die Angleichung eine transparente Etablierung des CG für die Probanden realisiert werden, die dazu führt, dass sie das adaptive System als zuverlässiger empfanden. Insgesamt lassen sich aus der Untersuchung Erkenntnisse der Anwendbarkeit des Interactive Alignment Modell (ebd.) auf die Mensch-Maschine-Kommunikation ableiten. So scheint auch der phonetisch-phonologischen Ebene der Angleichung eine bedeutsame Rolle zuzukommen, die in zukünftigen Untersuchungen Umsetzung finden sollte.

Auch diese Studie erweitert die Befundlage zu dem sognannten Konsistenzdilemma bei der Gestaltung von Dialogsystemen. Dabei zeigte sich erneut, dass inkonsistentes Systemverhalten keinen negativen Einfluss auf die Systembewertung hat.

Auch wenn die Ergebnisse nicht die positiven Effekte, die das Alignment in zwischenmenschlichen Kommunikationen aufweist, vollständig bestätigen konnte, so kann doch geschlossen werden, dass eine Übertragung zwischenmenschlicher Kommunikationsstrategien auf SDS als ein tragender Gestaltungsansatz gelten kann, wenn umfassende Dialogfunktionen durch die Umsetzung angemessen erfüllt werden.

Folgende Gestaltungsempfehlungen lassen sich aus der Untersuchung ableiten:

- Da die strategischen Komponenten im Mensch-Maschine-Dialog eine untergeordnete Rolle spielen, sollte die Angleichung stets als implizite Bestätigung fungieren und damit eine **Zeiteinsparung** ermöglichen (z. B. durch die Implementation eines Barge-In).
- Alignment sollte auf allen **linguistischen Ebenen** umgesetzt werden.
- Angleichende Systeme wurden als konsistenter wahrgenommen: **Konsistenz** im Rahmen der Gestaltung von SDS sollte nicht bedeuten, immer gleich zu reagieren, sondern nutzer- und situationsangemessen.
- Anhand der geäußerten Sprachausgaben antizipieren Nutzer die Fähigkeit des Systems Sprache zu verstehen und schätzen somit seine Fehleranfälligkeit. Ein variantenreiches **Promptdesign** vermittelt somit eine hohe Systemkompetenz.

Nachdem nun die drei Untersuchungen dokumentiert wurden, soll im folgenden Kapitel eine vergleichende Betrachtung der Studienergebnisse erfolgen.

6.4 Vergleichende Betrachtung der Studienergebnisse

In diesem Kapitel sollen die Ergebnisse der drei Hauptstudien und einer Vorstudie miteinander in Beziehung gesetzt werden. Auch wenn die Fragestellungen und Versuchspläne sich leicht unterschieden, so waren der Untersuchungskontext, die Stichprobeneigenschaften, die Erhebungsinstrumente, die Aufgabenstellung mit dem Sprachdialogsystem und die Fahraufgaben sehr ähnlich. Aufgrund der gemischten Versuchspläne und da bereits eine umfangreiche Analyse der Effekte innerhalb der Untersuchungen in den vorangegangenen Kapiteln stattgefunden hat, soll lediglich eine deskriptive Gegenüberstellung erfolgen.

Anfänglich soll eine intensive Betrachtung der globalen **Usability** geschehen. Hierzu wird der System Usability Score betrachtet, der in allen Untersuchungen unter nahezu gleichen Bedingungen erhoben wurde. Diese Analyse soll Hinweise darauf geben, welche zwischenmenschliche Kommunikationsstrategie das höchste Potential bei einem Transfer auf ein SDS aufweist. Darauffolgend soll eine Analyse von **Visualisierungsalternativen** erfolgen. Dazu werden die Umsetzungen der Studien I und II miteinander in Beziehung gesetzt. Da die Gestaltung auf Erkenntnissen der chronologisch früheren Studienergebnisse fußt, wird eine Verbesserung der Bewertung erwartet. Ebenfalls wurden die Blickabwendungen studienübergreifend betrachtet.

Nachdem Kernaspekte der Gestaltung von Sprachdialogsystemen aus der zusammenfassenden Betrachtung der Regressionsanalysen identifiziert und Gestaltungsempfehlungen ausgesprochen werden, erfolgte abschließend eine Diskussion der studienübergreifenden Ergebnisse und eine kritische Betrachtung des Untersuchungskontexts aller drei Studien.

6.4.1 Gebrauchstauglichkeitsbeurteilung

Zur zusammenfassenden Analyse des Einflusses der Dialoggestaltungsfaktoren auf die globale Usability-Bewertung erfolgte eine vergleichende Betrachtung des SUS-Scores aller Untersuchungen. Es ergaben sich vier Studien mit 15 Dialogbedingungen, die jeweils eine Stichprobenstärke zwischen fünf und 20 Versuchspersonen aufwiesen. Der Verlauf der SUS-Scores über die Erhebungen hinweg kann Abbildung 52 entnommen werden.

Die dynamische Dialogbedingung mit alleinigem Zustandsfeedback erlangte dabei den höchsten Wert der allgemeinen Gebrauchstauglichkeit, wohingegen die Einschätzungen der Vorstudie die geringsten Werte erzielten. Für alle drei Hauptstudien lag der SUS-Score über dem von Sauro (2011) angegebenen Durchschnittswert von 68. Die Einschätzung der Dialogbedingungen der Studie III entsprach ungefähr denen der Studie I. Nur die Kontrollgruppe der Studie I erlangte so hohe Werte auf der SUS Skala, dass sie nahezu mit den dynamischen Dialogbedingungen der Studie II gleich auf war.

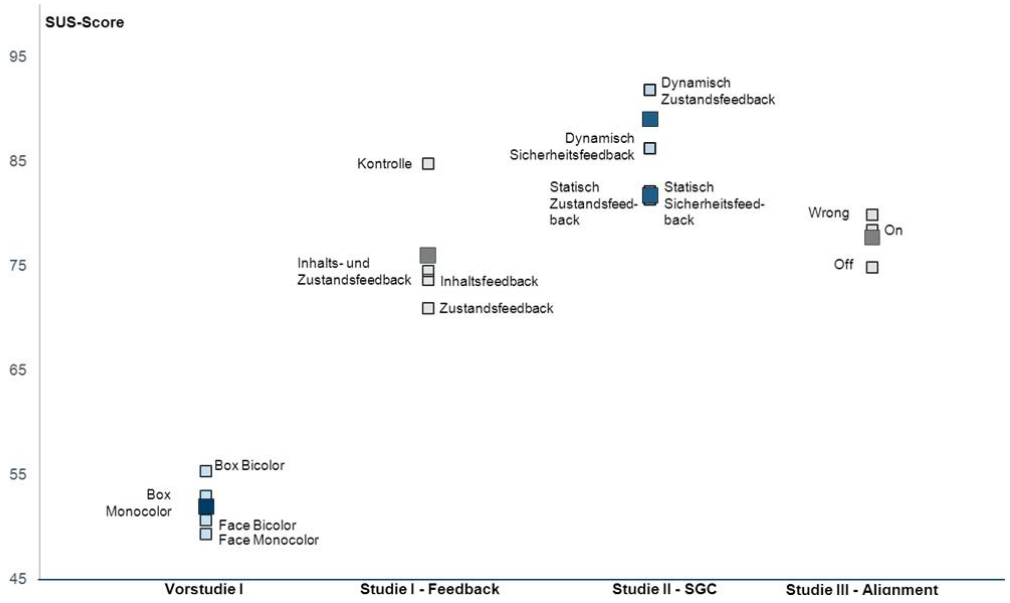

Abbildung 52: Studienübergreifender Verlauf der SUS-Scores

Um einen Hinweis darauf zu bekommen, welche der Manipulationen am meisten Potential zeigte, wurde die Betrachtung auf die Experimentalgruppen der Hauptstudien beschränkt. Es ergaben sich 6 Gruppen mit einer Gesamtstichprobe von *N*=75.

Es zeigte sich, dass sich insbesondere die dynamischen Dialogverhaltensweisen positiv von den visuellen Feedbackanzeigen und der Aligmentbedingung abhoben. Während beide Experimentalgruppen der Studie II SUS-Rohwerte von über 85 zeigten und damit besser als 90% der erreichten Scores eingestuft werden konnten (Sauro, 2011), lagen die Werte der anderen Dialogbedingungen unter 80.

Zur ergänzenden Betrachtung dieses Anstiegs der globalen Usability sollen die Bewertungen der Visualisierungen hinzugezogen und die gestalterischen Umsetzungen der grafischen Anzeigen miteinander verglichen werden

6.4.2 Visualisierungen

Um die Weiterentwicklung der Visualisierungen zu überprüfen, wurde der **Gestaltungskoeffizient** der Visualisierungsgruppen innerhalb der Studien I und II betrachtet. Dabei wurde sich bei der zweiten Hauptuntersuchung auf das statische System beschränkt, um die positiven Effekte der Flexibilisierung des Rückfrageverhaltens aus dem Vergleich auszuschließen.

In beiden Untersuchungen berechnete sich der Gestaltungskoeffizient aus den Bewertungen der Versuchspersonen, der Dimensionen Größe, Lesbarkeit, Attraktivität sowie einer allgemeinen Anzeigenbewertung. Abbildung 53 zeigt den Verlauf des Gestaltungskoeffizienten. Es ließ sich insbesondere ein Unterschied zwischen der Umsetzung des Zustandsfeedbacks und den anderen Visulisierungen beobachten. Während die alleinigen Rückmeldungen der Systemzustände als *mittel* bewertet wurden, wurden die anderen Umsetzungen der Studie I und II als *gut* bewertet.

Bei Betrachtung der einzelnen Bewertungsskalen fand man Unterschiede nach dem gleichen Muster. So wurden die Visualisierungen der zweiten Erhebung verständlicher und lesbarer als das Zustandsfeedback der ersten Erhebung bewertet.

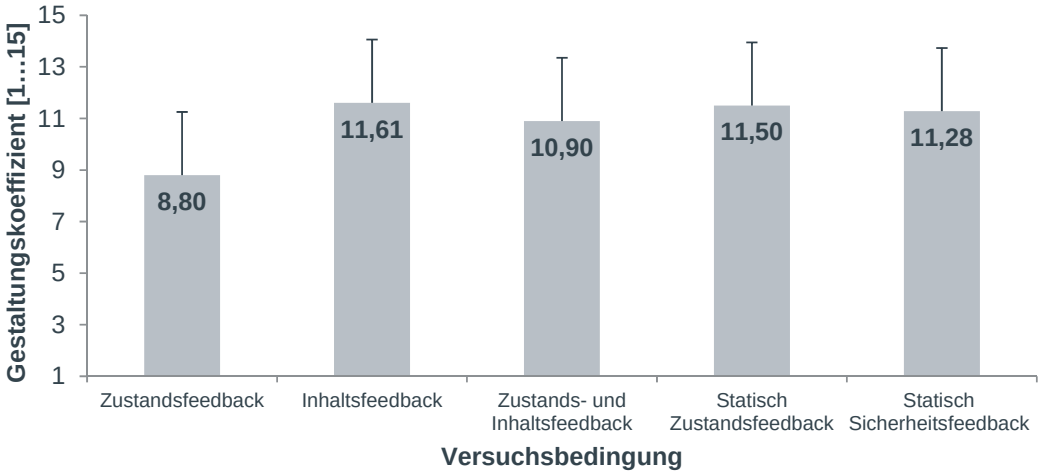

Abbildung 53: Studienübergreifender Verlauf der Gestaltungsbewertung (1 = sehr schlecht 15 = sehr gut) in Abhängigkeit der Versuchsbedingung

Für die fünf Versuchsgruppen, die sich aus den Studien I und II ergeben, wurden auch die **Blickabwendungen** vergleichend betrachtet. Die Werte für die prozentuale Häufigkeit der Kombiblicke in den Aufgabenfahrten lagen für 49 Versuchspersonen vor.

Dabei zeigte sich ein Anstieg der Blickabwendung über die Erhebungen (siehe Abbildung 36). Während innerhalb der Zustandsfeedbackgruppe (Studie I) weniger als 3% Kombiblicke vorgenommen wurden, waren es in der statischen Zustandsfeedbackgruppe (Studie II) über 13%.

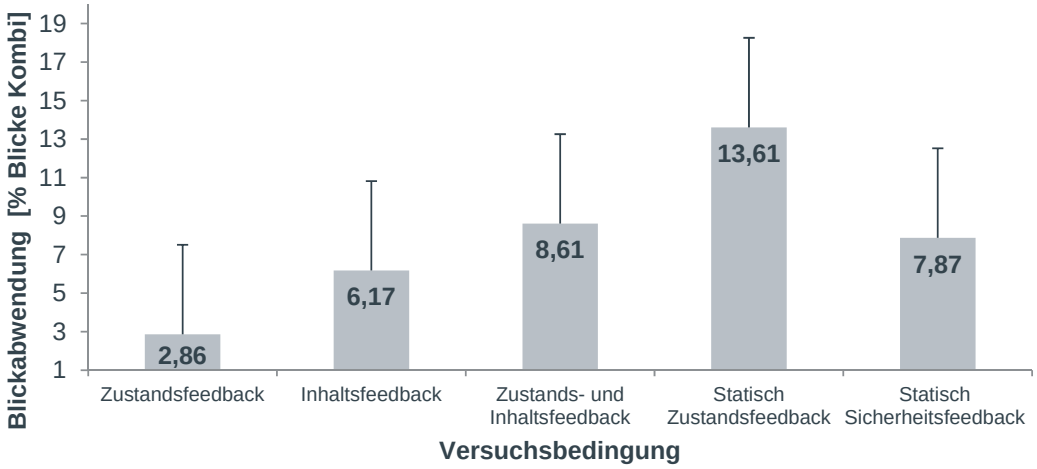

Abbildung 54: Studienübergreifender Verlauf der prozentualen Anzahl von Blicken auf das Kombidisplay in Abhängigkeit der Versuchsbedingung

6.4.3 Regressionen

Im Rahmen von übergreifenden (schrittweisen) Regressionsanalysen sollte identifiziert werden, welche Parameter die Usabilityeinschätzung von Sprachdialogsystemen beeinflussen. Als Regressand wurde der SUS-Score gewählt. Da sowohl der SUS-Score, als auch die erklärenden Variablen in gemischten Versuchsdesigns erhoben wurden, sollten die folgenden Analysen eher deskriptiv verstanden werden.

Zunächst sollte überprüft werden, welchen Einfluss die **objektiven Parameter** der Effizienz (Dialoglänge in Sekunden) und der Effektivität (FE und TTE) auf die zu erklärende Größe haben. Die Aufklärung der Varianz kann als unabhängig von dem Methoden Bias angesehen werden, da die Dialogdauer aufgezeichnet wurde und die Fehlbedienungen und -erkennungen protokolliert vorlagen. Bei der Modellerstellung wurde der SUS-Gesamtscore allein durch die Anzahl der Fehlerkennungen (β= -.32, p< .01) vorhergesagt (korrigiertes R^2= .10).

Zur Analyse sollten nun auch die **subjektiven Beurteilungen** der Probanden bezüglich der Effizienz, Effektivität und Zufriedenheit hinzugezogen werden. Als Regressoren ging neben der Dialoglänge, der Anzahl der TTE und FE auch die Anzahl der Turn Takes, die Anzahl der erfolgreichen Dialogabschlüsse, die Angemessenheitsbeurteilung der Dialoglänge und die

Zufriedenheitsbekundung ein[35]. Da diese Variablen nur in der Studie II und III in ähnlicher Weise erhoben wurden, reduzierte sich der Datenpool auf die beiden letzten Erhebungen.

Die multiple, schrittweise Regression zeigte, dass der SUS-Score durch die Zufriedenheitsbeurteilung und die Anzahl der Turn Takes vorhergesagt werden konnte (Tabelle 38; korrigiertes R^2= .50).

Tabelle 38: Multiple, schrittweise Regression zur Vorhersage SUS

			Effekt SUS
	β	***T***	***p***
Zufriedenheit	**.68**	**9.6**	**< .0001**
Turn Takes	**-.14**	**-2.0**	**.05**

Darauf aufbauend sollten anhand der objektiven Parameter (Dialoglänge, Anzahl TurnTakes, Erfolg, FE, TTE) die **Stellgrößen der Zuverlässigkeit** identifiziert werden. Bei der Betrachtung der Zuverlässigkeit konnten die Anzahl der Fehlerkennungen (β= -.39, p< .01) und die Anzahl der TTE (β= .28, p< .01) als signifikante Regressoren identifiziert werden (korrigiertes R^2= .27). Als **Stellgrößen der Zufriedenheit** konnten wiederrum die Zuverlässigkeit (β= .61, p< .01) und die Angemessenheit der Dialoglänge (β= .34, p< .01) identifiziert werden (korrigiertes R^2= .70).

Zur Vorhersage der **Blickabwendungen** wurde über die Studien hinweg auch eine multiple Regression gerechnet. Als Regressoren gingen die Fehlerkennungen und -bedienungen, sowie der Gestaltungskoeffizient und der SUS-Score in die schrittweise Analyse mit ein. Es zeigte sich, dass die Blickabwendung signifikant durch die Anzahl der TTE-Fehler (β= -.29, p= .02) vorhergesagt werden kann (korrigiertes R^2= .07).

[35] Die Zuverlässigkeitseinschätzung musste aus den Analysen ausgeschlossen werden, da eine zu hohe Interkorreliertheit der Regressoren vorlag.

6.4.4 Diskussion

In diesem Kapitel sollen die Ergebnisse der vergleichenden Betrachtung und der allgemeine Untersuchungskontexts der drei Studien diskutiert werden.

6.4.4.1 Diskussion der Ergebnisse

Über die Untersuchungen hinweg sollte zunächst die Qualität und die Entwicklung der **Gebrauchstauglichkeit** betrachtet werden. Betrachtet man die Werte der SUS über die Studien hinweg, so zeigte sich, dass sich die dynamischen Feedbackgruppen der Studie positiv von den anderen Gruppen abheben. Eine Ausnahme bildete die Kontrollgruppe der Studie I. Zu dieser Gruppe, die keine Manipulation des ursprünglichen Dialogverhaltens des SDS beinhaltete, zeigten sich keine signifikanten Unterschiede. Dennoch erlang das dynamische Rückfrageverhalten, unabhängig von dem präsentierten Feedback, bessere Usabilitybewertungen als alle anderen Experimentalgruppen. Von allen Manipulationen des Dialogverhaltens, die immer eine zwischenmenschliche Kommunikationsstrategie auf ein SDS applizierten, scheint die Flexibilisierung des SGC nach dem LCE-Modell der vielversprechendste Ansatz. So unterscheiden sich die dynamischen Dialogbedingungen positiv von den visuellen Feedbackanzeigen und der Aligmentbedingung. Es kann geschlossen werden, dass der Flexibilisierung des SGC das größte Potential bei der nutzerorientierten Verbesserung der Gebrauchstauglichkeit von SDS zukommt.

Bei der Betrachtung der allgemeinen Gebrauchstauglichkeit muss jedoch kritisch betrachtet werden, dass im Verlauf der Studien Verbesserungen der Effektivität des Sprachdialogsystems vorgenommen wurden. Während in der Vorstudie und in der Studie I ein hohes Maß an Fehlerkennungen vorlag, konnte dies in Studie II und III stark reduziert werden. Davon profitierte insbesondere die Einschätzung der globalen Gebrauchstauglichkeit. Die Werte des SUS müssen demnach als konfundiert durch die Effektivitätssteigerung betrachtet werden.

Dass der Effekt des dynamischen Dialogverhaltens nicht allein durch die Verbesserung der Effektivität des Systems über die Studien hinweg erklärt werden kann, belegt der SUS-Gesamtwert der Studie III, der durchschnittlich viel geringer war als der Wert der Studie II. Vergleicht man die Werte der Studie II (statisch) mit denen der Studie III, so scheinen die Visualisierungen doch einen positiven Effekt auf die Usability des SDS zu haben. Dabei entsprechen die Einschätzung der Dialogbedingungen der Studie III ungefähr denen der Studie I. Den Visualisierungen kommt somit ungefähr das gleiche Potential zu, wie den Gestaltungen der Systemausgaben.

Durch das Untersuchungsdesign der zweiten Studie, welches das Rückfrageverhalten als Messwiederholungsfaktor betrachtete, können jedoch nicht nur die Effekte des Zwischensubjektfaktors Feedback unterschätzt werden (siehe 6.2.4.3), sondern auch die Bewertung des statischen Rück-

frageverhaltens. So ist es denkbar, dass durch die Kontrastierung zu dem dynamischen Feedbackverhalten die Bewertung der Usability des statischen Rückfrageverhaltens insgesamt verschlechtert wurde.

Neben einer deutlichen Verbesserung der Gebrauchstauglichkeit von der Vorstudie zu der ersten Hauptstudie, fanden sich bei Betrachtung des SUS-Scores der Visualisierungsgruppen auch Verbesserungen der Zustandsanzeigen von Studie I zu Studie II. Bei Studie II traten Zustands- und Inhaltsfeedback stets kombiniert auf. Dennoch ließ sich eine Steigerung der Gebrauchstauglichkeitsbewertung von der alleinigen Darbietung des ZF oder IF in Studie I zu der kombinierten Darbietung in Studie II feststellen, während dieser Anstieg innerhalb der Studie I nicht so präsent auftrat. Zu der kombinierten Feedbackdarstellung der Studie I zeigte die Umsetzung der Studie II einen leichten Anstieg. Insgesamt bestätigen die Ergebnisse die Schlussfolgerungen aus der Studie I, dass Inhalts- und Zustandsfeedback stets in Ergänzung zueinander dargeboten werden sollten.

Die Betrachtung des **Gestaltungskoeffizienten** zeigte eine konsistente Befundlage, wie der SUS-Score. Wobei sich hier die Anzeigen der Studie II lediglich von der Zustandsanzeige der Studie I positiv abheben konnten. Einen Unterschied zu den textuellen Anzeigen der Studie II ließ sich nicht feststellen. Dieser Unterschied des Gestaltungskoeffizienten zeigte sich sowohl auf die Verständlichkeit als auch auf die Größe und Lesbarkeit zurückführen. So waren die Anzeigen der Studie II verständlicher und lesbarer als das ZF der Studie I.

Abseits des Zustandsfeedbacks ließ sich keine Verbesserung des Gestaltungskoeffizienten ausmachen. Dies kann vor allem durch den Bezugsrahmen der Probanden begründet werden. So konnte beobachtet werden, dass die Anzeigen-spezifischen Beurteilungen über die Studien hinweg sehr ähnlich sind. Ohne eine explizite verbale Verortung der Pole konnte verzeichnet werden, dass Versuchspersonen gerade in einer Interviewsituation dazu tendieren, Anzeigen als *gut* zu bewerten. Über die Studien hinweg muss die Trennschärfe der Anzeigenbeurteilung als gering eingestuft werden. Da für die Probanden keine Möglichkeit bestand, die Umsetzungen des Inhaltsfeedbacks mit den Visualisierungen der Studie II zu vergleichen, lagen sie alle im selben Bewertungsspektrum. Eine Kontrastierung oder stringente Verortung der Pole wäre für folgende Erhebungen empfehlenswert.

Der Anstieg der **Blickabwendung** lässt sich einerseits dadurch erklären, dass dem Zustandsfeedback insgesamt wenig Aufmerksamkeit zuteilwurde. In Kapitel 6.1.4.1 wurde argumentiert, dass die Probanden die Anzeige häufig nicht beachteten. Dagegen lässt sich der Anstieg zu Studie II maßgeblich durch das veränderte Farbkonzept erklären. Insbesondere in der grauen Bedingung war die Unterscheidung zwischen einem grauen Mikrofon-Icon und einem ebenfalls grauen Lautsprecher-Icon auf grauem Hintergrund nur schwer peripher wahrnehmbar. In der sta-

tisch bunten Bedingung konnte eine Angleichung an das Niveau der kombinierten Feedbackvisualisierung der Studie I beobachtet werden. Die reduzierte Textdarbietung hatte damit zwar nicht den erwarteten Effekt die Blickabwendung zu verringern, zeigte damit aber auch keine negativen Auswirkungen. Die Befundlage spricht dafür, dass es hinsichtlich der Blickzuwendung zur Visualisierung keinen Unterschied macht, ob die Nutzerein- bzw. Systemausgabe als gesamter Text inklusive Füllworte dargestellt wird oder nur die Schlüsselworte angezeigt werden. Auch wenn empfohlen wird die Anzeige der Nutzereingabe und der Systemausgabe zu kombinieren um somit eine Verknappung des Inhaltsfeedbacks zu erreichen, so sollten anschließende Untersuchungen sich dieser Fragestellung erneut annehmen. Von zentraler Bedeutung erscheint die farbige Rückmeldung der Zustände, um die periphere Wahrnehmbarkeit zu begünstigen.

Interessant erscheint der Befund, dass im Rahmen der Regressionen die Blickabwendung (wenn auch nur mit einer geringen Varianzaufklärung) durch die Anzahl der TTE-Fehler vorhergesagt werden konnte. Auch wenn die Blickabwendung im Fahrtkontext und im Rahmen der vorliegenden Studien von sehr viel mehr Faktoren beeinflusst wurde, als von den Fehlbedienungen, so stützt dieser Befund doch die Annahme, dass unerwartetes Systemverhalten oder problematische Interaktionen zu Blickabwendungen führen.

Bei der globalen Betrachtung der regressiven Zusammenhänge wurde der **SUS**-Gesamtscore allein durch die Anzahl der Fehlerkennungen vorhergesagt (korrigiertes R^2=.10). Die geringe Varianzaufklärung lässt sich vor allem dadurch erklären, dass nur objektive Parameter der Usability als Regressoren in die Analyse mit einbezogen wurden. Diese zeigten keine methodische Ähnlichkeit zur Erhebung des SUS-Scores über einen Fragebogen. Umso beachtlicher ist, dass die Anzahl der protokollierten Fehlerkennungen die globale Usability Bewertung der Probanden vorhersagen konnte. Zieht man die subjektiven Parameter hinzu, so bestand nicht nur zwischen den Regressoren eine Multikollinearität, sondern auch ein gewisser Methoden Bias. Insbesondere die Bewertung der Zuverlässigkeit des Systems, der Angemessenheit der Dialoglänge und der Zufriedenheit war der Erhebungsart des SUS in allen Studien sehr ähnlich. Die Zuverlässigkeitsbeurteilung wurde aufgrund der zu hohen Interkorreliertheit mit der Zufriedenheit aus den Analysen ausgeschlossen. Es zeigte sich, dass der SUS-Score nun durch die Zufriedenheit und die Anzahl der Turn Takes vorhergesagt werden kann. Wobei weiterführend die Zuverlässigkeit und die Angemessenheit der Dialoglänge die Zufriedenheit vorhersagen konnten. Die Zuverlässigkeitsbeurteilung wurde ihrerseits vor allem durch die Anzahl der Fehlerkennung erklärt. Die Interaktionslänge und die Zuverlässigkeit scheinen insgesamt einen bedeutsamen Einfluss auf die Gebrauchstauglichkeit von SDS zu haben. Dies bietet erneut einen Hinweis darauf, dass bei gegebener Effektivität insbesondere die Effizienz bei der Interaktion mit einem SDS eine maßgebliche Rolle spielt. Erneut sei daraufhin hingewiesen, dass dieser Befund in gewisser Weise durch die Manipulation des SGC in Studie II beeinflusst ist.

6.4.4.2 Diskussion Methode

In allen drei Untersuchungen wurden die Systeme von **Mitarbeitern des Volkswagen Konzerns** bewertet. Obwohl die Varianz der Tätigkeiten am Standort Wolfsburg als hoch einzustufen ist, kann davon ausgegangen werden, dass die Stichprobe eine höhere Technik- und Automobilaffinität besaß als die Grundgesamtheit. Durch die Rekrutierung über eine betriebsinterne Probandendatenbank, kann ein gewisser Grad an Selbstselektion und ein Sponsorship-Effekt nicht ausgeschlossen werden. Weiterführende Untersuchungen sollten mit Stichproben erfolgen, die die Zielgruppe des Systems besser abbilden.

Weiterhin können die Erkenntnisse der Experimente nicht über den **europäischen Kulturkreis** hinaus generalisiert werden. Insbesondere die Ergebnisse der Studie II, die die Verbesserung der Usability von SDS durch die Flexibilisierung des Rückfrageverhaltens im Sinne des Least Collaborative Effort Ansatz belegen konnten, sollten in internationalen Studien gefestigt werden. Da Hamerich (2009) und Leiber (2010) berichten, dass in Japan und in China eine Präferenz für lange und ausführliche Systemausgaben besteht, müssten fortführende Studien untersuchen, ob die dynamische und effizientere Dialoggestaltung auch in den asiatischen Kulturen zu einer Steigerung der Nutzerzufriedenheit führen kann. Weiterhin sollten die Ergebnisse stets im Rahmen des **Fahrtkontextes** betrachtet werden, in dem die Interaktion mit dem SDS lediglich eine Zweitaufgabe darstellt, für die der Fahrer nicht alle Aufmerksamkeitsressourcen aufwenden kann. Eine Generalisierbarkeit der Befunde auf sprachverarbeitende Systeme in Auskunfts- oder Mobilfunkkontexten ist nicht unbedingt gegeben.

Im Rahmen der Untersuchungen wurde immer wieder deutlich, dass es sich bei Sprache um eine hochgradig **flüchtige Modalität** handelt. Eine Bewertung und Verbalisierung der Dialogführung gestaltete sich für die Versuchspersonen schwierig. Zumeist zogen sie als Bewertungsheuristik negative Erfahrungen (wie z. B. Fehlerkennungen) heran, ohne sich an den konkreten Dialogablauf erinnern zu können. Nur starke Abweichungen von der Dialogeffektivität oder -effizienz wirkten sich auf die Systembeurteilungen negativ aus. In zukünftigen Untersuchungen sollte daher geprüft werden, ob sich die Erhebung erleichtern lässt, wenn direkt nach jeder Systeminteraktion nach der mentalen Belastung und der Dialoggestaltung gefragt wird. Auch das Einbeziehen einer Videoanalyse scheint ein vielversprechender Untersuchungsansatz, um Versuchspersonen den Dialogablauf anhand ihres eigenen Interaktionsvideos bewerten zu lassen.

Weiterhin fußten alle Untersuchungen auf einer vergleichsweise kurzen Interaktionsdauer mit den Sprachdialogsystemen. **Langzeituntersuchungen**, welche Nutzungshäufigkeiten und Gewöhnungseffekte genauer abbilden können, sollten angeschlossen werden. Insbesondere die lexikalischen Angleichungen der Studie III könnten von einer längeren Nutzungsdauer profitieren.

Ferner muss geschlussfolgert werden, dass die Etablierung der zwischenmenschlichen Strategien den Dialog mit dem SDS nur bedingt in Richtung automatische Prozesse verschieben konnte. In keiner der Studien ließ sich ein positiver Effekt der Dialogprinzipien auf die mentale Beanspruchung feststellen. Auf der anderen Seite führen Interaktionen mit einem SDS nur zur einer geringen mentalen Belastung. Weitere Reduktionen durch das Dialogdesign sind deshalb nur in geringem Ausmaß möglich. Um diese diffizilen Differenzen zu erfassen, sollte die Messung des **Automatisierungsgrads** der Dialogführung feingranularer erfolgen. Innerhalb der Simulatorfahrten mussten die Probanden lediglich die Längs- und Querführung regeln, während in normalen Verkehrssituationen viele weitere beanspruchende Faktoren hinzukommen (z. B. Navigieren, andere Verkehrsteilnehmer). Auch Ranney (2008) konnte belegen, dass das Fahren auf realen Straßen mehr von Zweitaufgaben beeinflusst wird, als das Fahren in einem Simulator. Zur besseren Erfassung der kognitiven Beanspruchungseffekte des SDS auf die Fahreraufmerksamkeit sollten in Zukunft Reaktionstests (wie der peripheral detection task (PDT)) in Ergänzung zu den üblichen Fahrparametern in Betracht gezogen werden (siehe Kap. 6.3.4.2).

Trotz des hohen Standardisierungsgrads, der durch das **Vorsprechen** der Kommandos erreicht werden konnte, sollten zukünftige Studien um einen Untersuchungsteil ergänzt werden, in denen Versuchspersonen das SDS frei explorieren können. Das Vorsprechen der Kommandos kann gerade die kognitiven Ressourcen unterschätzen, die zur Konzeptualisierung und Formulierung einer Sprachäußerung notwendig sind und Fehlerraten künstlich reduzieren. Eine realistischere Abbildung der Bedienung sollte bei freierer Vorgabe der Interaktionsziele in zukünftigen Erhebungen angestrebt werden.

Innerhalb des Messwiederholungsdesigns kann eine Alternative zur Präferenzbekundung das **no-choice/choice Paradigma** sein, welches bei Brumby et al. (2011) ausführlich beschrieben ist. Neben einer objektiven Messung der einzelnen Bedingungen ermöglicht es die Identifikation der präferierten Option durch die Nutzer und kann aufzeigen, ob sich diese Bedingungen unterscheiden.

Die Verwendung Fragestellung-spezifischer Items mit einer **Kategorienunterteilungsskala** erwies sich jedoch als zielführend. Insbesondere die Befragung im Rahmen von Interviews konnte durch eine Verankerung der einzelnen Pole profitieren. Ebenfalls wies der SUS Fragebogen der zur Feststellung der allgemeinen Gebrauchstauglichkeit in nahezu allen Studien erhoben wurde, eine hohe interne Konsistenz auf. Der Vergleich zu anderen, parallel erhobenen Messinstrumenten konnte ebenfalls eine hohe Validität belegen. Aktuelle Arbeiten (vgl. Lewis & Sauro, 2009), die die Eindimensionalität der SUS Skala hinterfragen, konnten nicht bestätigt werden.

Zusammenfassend sollen Gestaltungsempfehlungen aus den Erkenntnissen der drei Studien und ihrer vergleichenden Betrachtung abgeleitet werden.

6.5 Fazit

Die Darstellung der **Dialoginhalte und Systemzustände** der ersten Erhebung konnte die Dialogführung nicht in dem Ausmaß unterstützen, wie es zu erwarten gewesen wäre. Die Untersuchung konnte allerdings belegen, dass die Gestaltung der Anzeigen eine zentrale Rolle spielt. Insbesondere die Größe und das Timing der visuellen Anzeigen sind sensible Umsetzungsparameter, die nicht verletzt werden dürfen. Die Erkenntnisse der vergleichenden Betrachtung deuten allerdings darauf hin, dass die Größe der Anzeige aus Studie II als ausreichend bewertet werden kann.

Die kombinierte Feedbackdarstellung konnte dabei zufriedenstellende Gebrauchstauglichkeitswerte erzielen und die Einschränkungen der alleinigen Feedbackdarstellungen ausgleichen. Es kann demnach geschlossen werden, dass eine Zustandsfeedbackanzeige nur in Kombination mit Inhaltsfeedback erfolgen sollte. Der studienübergreifende Vergleich konnte belegen, dass es kaum einen Unterschied macht, ob die Dialoginhalte mit oder ohne Füllwort angezeigt werden. Weiterhin deuten die Ergebnisse darauf hin, dass die farbige Rückmeldung der Zustände die periphere Wahrnehmung begünstigen und somit die Blickabwendung reduzieren kann.

Bei der Umsetzung von Inhalts- und Zustandsfeedback sollte sich daher an den farbigen Varianten der Studie II orientiert werden. Die abschließende vergleichende Betrachtung zeigt, dass die Visualisierungsgruppen der Studie II eine höhere Gebrauchstauglichkeit aufweisen als die Gruppen der Studie III, die keine Visualisierungen zeigten. Somit scheinen die Visualisierungen, bei gegebener Effektivität, doch einen leicht positiven Effekt auf die Usability des SDS zu haben.

Neben den grafischen Rückmeldekonzepten erscheint die Umsetzung eines **dynamischen Rückfrageverhaltens** als eine Weiterentwicklung aktueller Sprachdialogsysteme, die mit einer hohen Priorität versehen ist. So ermöglichte die Flexibilisierung des System Grounding Criterion in Studie II eine Steigerung der Gebrauchstauglichkeit und erhöhte die nutzerseitige Zufriedenheit. In nahezu allen Untersuchungen konnte belegt werden, dass der Effizienz eine bedeutende Rolle im Mensch-Maschine-Dialog zukommt. Insbesondere im Rahmen der zielorientierten Dialoge und der mentalen Belastungssituation des Fahrtkontextes scheint eine effiziente Dialogführung essentiell. Selbst Fehlauslösungen wurden nutzerseitig akzeptiert, sofern die Mehrzahl der Dialoge dadurch zügiger vollzogen werden konnte.

Die Definition des SGC über die **Erkennerkonfidenz** stellte sich dabei als vielversprechend heraus. Zur umfassenden Integration des Konzepts in alle sprachbedienbaren Kontexte des automotiven HMIs sollten jedoch Anpassungen des dynamischen Dialogverhaltens an die Kritikalität der Aufgabe erfolgen. Die farbige Vermittlung der Erkennersicherheit sollte dabei stets über eine gut verständliche Metapher erfolgen. Die Verwendung eines Farbcodes in Form einer Ampel erwies

sich als erfolgreich und konnte das dynamische Verhalten des Systems transparenter gestalten. Die farbige Rückmeldung der Systemzustände begünstigte dabei sogar die periphere Wahrnehmbarkeit und konnte erneut die Blickabwendungen minimieren. Die Reihenfolgeeffekte der zweiten Erhebung können belegen, dass die dynamische Anpassung der Interpretationsstrategien von Beginn an geschehen sollte. Eine Adaption im Nutzungsverlauf von statischem zu dynamischen Rückfrageverhalten verwirrte die Nutzer und führte zu einer verringerten Akzeptanz.

Die Erkenntnisse der Studie II verdeutlichen darüber hinaus, dass den visuellen Rückmeldungen lediglich ein ergänzender Charakter zukommen sollte. Die akustische Darbietung sollte stets alle Informationseinheiten bei Aktionsankündigung beinhalten. Die Nutzer reagierten gerade zu Beginn der Interaktion sehr sensibel auf die reduzierte, akustische Darbietung.

Die Umsetzung des **linguistischen Alignments** in Anlehnung an das Interactive Alignment Modell (Pickering & Garrod, 2004) konnte dagegen nur bedingt eine Erleichterung des Grounding-Prozesses mit einem Sprachdialogsystem ermöglichen und die Gebrauchstauglichkeit nicht steigern. Ausschließlich Nutzer mit SDS-Erfahrung bewerteten das System, welches eine lexikalische Spiegelung der Nutzereingabe zeigte, positiver. Novizen bewerteten dagegen eine hohe Variabilität der Systemausgaben unabhängig von ihrer eigenen Formulierung als positiv. Für spätere Umsetzungen sollte die Zeiteinsparung, die Alignmentprozesse ermöglichen, mit betrachtet werden. In dem die Angleichung als implizite Bestätigung fungiert und Nutzer das System bei gegenseitigem Verständnis unterbrechen können, sollten die effizienzbasierten Vorteile des Alignments stärker zum Tragen kommen. Aufgrund der Befundlage angrenzender Forschungsarbeiten kann davon ausgegangen werden, dass sich die Implementation von systemseitigen Alignment vor allem in Problemlösekontexten und Tutoriumssituationen (Hilfekontexte, Bordbücher) lohnt. Eine weitere Empfehlung wäre, die Angleichung auf allen linguistischen Ebenen zu implementieren. Neben der lexikalischen und syntaktischen, sollte auch die phonologische Ebene der Eingabe gespiegelt werden. Dadurch sollten die positiven Einflüsse der unvermittelten Priming-Prozesse auf die Fahreraufmerksamkeit verstärkt werden.

Sowohl in Studie II, als auch in Studie III wurden adaptive Systeme evaluiert, welche eine hohe Dynamik in den Systemreaktionen zeigten. In beiden Studien konnte belegt werden, dass sich dies nicht negativ auf die **Konsistenzbewertungen** auswirkte. Es kann geschlussfolgert werden, dass Konsistenz im Rahmen der Gestaltung von SDS demnach nicht bedeutet, dass das System immer gleich reagiert, sondern dass es eher nutzer- und situationsangemessene Reaktionen zeigen sollte. Nachdem nun umfassende Gestaltungsempfehlungen abgeleitet wurden, schließt die Arbeit mit einer allgemeinen Diskussion.

7 Zusammenfassende Diskussion

Dieses Kapitel soll die vorliegende Arbeit mit einer zusammenfassenden Betrachtung abschließen und die Frage beantworten, ob eine Übertragung zwischenmenschlicher Kommunikationsprinzipien auf SDS erfolgsversprechend ist.

Im ersten Teil der Arbeit wurden die Grundlagen, Potentiale und Einschränkungen aktueller Sprachdialogsysteme dargestellt. Die Collaborative Theory wurde herangezogen, um die geringe Akzeptanz derzeitiger SDS zu begründen und Optimierungsbereiche aufzuzeigen. Anschließend wurden Ansätze zur Verbesserung der Gebrauchstauglichkeit aus den zwischenmenschlichen Kommunikationsprinzipien abgeleitet. In Bezug auf Feedbackprozesse, Interpretationsstrategie und Promptdesign wurde sich an der kollaborativen Beziehung menschlicher Gesprächspartner orientiert, um ein bestehendes automotives SDS weiterzuentwickeln.

Im zweiten Teil der Arbeit wurden diese Umsetzungen in experimentellen Nutzerstudien evaluiert und hinsichtlich ihrer theoretischen und praktischen Konsequenzen eingeordnet. Im Speziellen wurden die Integration visueller Rückmeldeprozesse, die Flexibilisierung der Interpretationsstrategie und die lexikalische Spiegelung des Nutzerinputs auf ihre Transferierbarkeit aus der zwischenmenschlichen Kommunikation in den Mensch-Maschine-Dialog überprüft. Nachdem umfangreiche Gestaltungsempfehlungen aus den Ergebnissen abgeleitet wurden, soll Kapitel 7.1 nun eine Einordnung der Arbeit in den Gesamtzusammenhang der Forschungslage bieten. Anschließend sollen Anstöße für fortführende Optimierungsbereiche aufgezeigt werden (Kap. 7.2). Die Arbeit endet mit einem Ausblick.

7.1 Einbettung in bestehende Befundlage

Die vorliegende Arbeit adressierte die Fragestellung, ob sich die zwischenmenschlichen Kommunikationsstrategien der Collaborative Theory (Clark, 1996) als Gestaltungsansatz für das Dialogdesign eines automotiven SDS eignen. Dabei konnte die bestehende Forschungslage um Erkenntnisse ergänzt werden, die in experimentellen Nutzerstudien gewonnen wurden. Die Anpassung des Systems an bestehende kollaborative Kommunikationsstrukturen wurden somit nicht nur als Gestaltungsmaxime vorgeschlagen (z. B. Brennan & Hulteen, 1995), sondern in standardisierten Bedingungen auf ihren Einfluss auf die Gebrauchstauglichkeit eines automotiven Systems untersucht. Somit können Aussagen darüber abgeleitet werden, welche Bestandteile der Collaborative Theory die Grounding-Prozesse zwischen Mensch und Maschine erleichtern und den Bedienablauf nutzerfreundlicher gestalten können.

Im Rahmen der ersten Erhebung konnten Ergebnisse zum Forschungsstand beigetragen werden, die die Fragestellung adressieren, wie sich grafische **Zustands- und Inhaltsanzeigen** eines au-

tomotiven Sprachdialogsystems auf die nutzerseitige Systembewertung und Interferenz mit der primären Fahraufgabe auswirken. Der Übertragung beider Dialogspuren (Clark, 1996) auf den Mensch-Maschine-Dialog kam bei den getesteten Umsetzungen nicht die erwartete Bedeutung zu. Dennoch konnte belegt werden, dass Nutzer Visualisierungen als dialogstrukturierende Signale verwenden und sich eine derartige Visualisierung wünschen. Der ausschließlich ergänzende Charakter der Visualisierungen muss allerdings hinterfragt werden. Solange sie dem Nutzer keine erweiterte Handlungsfähigkeit bieten, werden sie ihrer Rolle in der zwischenmenschlichen Kommunikation nicht gerecht und ihr Mehrwert muss als gering eingestuft werden. Die Fragestellung, welche Bedeutung Feedbacksignalen in der Interaktion mit einem SDS zukommt, sollte mit einer verbesserten Umsetzung und umfangreicherer Funktionsanbindung (z. B. Korrekturmöglichkeiten im Rahmen des Inhaltsfeedbacks) erneut untersucht werden. Eine klare Gestaltungsempfehlung lautet, dass Statusanzeigen stets mit einer Erweiterung der Funktionalität (z. B. Barge-In) einhergehen sollten, um die Dialogeffizienz durch die Rückmeldungen zu steigern.

Durch die dritte Untersuchung konnte die bestehende Befundlage um Erkenntnisse ergänzt werden, wie sich die **lexikalische und syntaktische Angleichung** eines automotiven Sprachdialogsystems auf die nutzerseitige Systembewertung auswirken. Ohne eine gleichzeitige Verringerung des zeitlichen Aufwandes durch das Alignment konnte dabei keine Steigerung der allgemeinen Gebrauchstauglichkeit nachgewiesen werden. Die strategischen Vorteile des Alignments konnten nur innerhalb einer Expertenstichprobe beobachtet werden.

Während van Baaren et al. (2003) eindrucksvoll belegen konnten, dass Alignment im zwischenmenschlichen Kontext sogar zu monetären Begünstigungen führen kann, reagierten nicht alle Nutzer positiv auf die lexikalische Spiegelung des Systems. Obwohl auch bei der Kellnerin (ebd.), die jedes Element der Bestellung wiederholte, keine Zeiteinsparung durch das Alignment beobachtet werden konnte, verstärkt sie durch ihr Verhalten die Bindung zu ihren Gästen und wird durch ein erhöhtes Trinkgeld belohnt. Nutzer eines automotiven SDS scheinen allerdings kein Interesse daran zu haben, affektive Bindungen zu Maschinen einzugehen, insbesondere wenn sie sich in mental beanspruchenden Situationen befinden. Im Mensch-Maschine-Dialog im Fahrtkontext lassen sich daher kaum strategische Vorteile des linguistischen Alignments aufzeigen.

Weiterhin können aus der Untersuchung auch Erkenntnisse der Anwendbarkeit des Interactive Alignment Modells (Pickering & Garrod, 2004) auf die Mensch-Maschine-Kommunikation abgeleitet werden. Die Ergebnisse lassen vermuten, dass Alignment auf Ebenen, die nicht bedeutungsrelevant sind (z. B. Betonung), ebenfalls eine fundamentale Rolle in der Angleichung semantischer Repräsentationen spielen. So scheint auch der phonetisch-phonologischen Ebene der Angleichung eine bedeutsame Rolle zuzukommen. Eine Umsetzung des Alignments auf allen Ebenen des Modells sollte Aufschluss darüber geben, welche Bedeutung der prosodisch-phonologischen Spiegelung bei der mentalen Entlastung des Fahrers im Dialog mit dem SDS

zukommt. Eine Umsetzung des Alignments sei daher nur auf allen Ebenen des Interactive Alignment Modells (ebd.), bei Funktionserweiterung mit einhergehender Zeiteinsparung und in Kontexten mit Tutoriumscharakter angeraten.

Dass die Zeiteinsparung als eine der bedeutendsten Stellschrauben der Gebrauchstauglichkeit eines SDS gelten kann, belegte auch die zweite Untersuchung. Am Beispiel der **Flexibilisierung** des Interaktionsprozesses konnte nachgewiesen werden, dass Dialoggestaltung im Sinne des kleinsten gemeinsamen Aufwands (Least Collaborative Effort) zur Verbesserung des Mensch-Maschine-Dialogs beitragen kann. Somit konnte ohne eine technische Verbesserung der Erkennergüte, eine Steigerung der Bedienbarkeit und Akzeptanz aktueller Sprachdialogsysteme erreicht werden. Die Studie konnte belegen, dass sich die Einsparung überflüssiger Bestätigungsanfragen auf allen Facetten der Usability positiv auswirkt. Die Umsetzung einer flexiblen Interpretationsstrategie wird somit kontextübergreifend angeraten.

Im Rahmen der Erhebung konnte sogar beobachtet werden, dass Probanden Fehler des Systems für eine höhere Effizienz akzeptierten. Je länger eine Interaktion dauerte, umso geringer wurde die Fehlertoleranz der Nutzer. Dieser Befund steht in Einklang mit den Erkenntnissen von Brumby et al. (2011), Wechsung et al. (2010) und Cameron (2010), die ebenso der Anzahl der Interaktionsschritten eine große Bedeutung bei der nutzerseitigen Akzeptanz beimessen.

Insbesondere im Fahrtkontext, in dem den Nutzern nur begrenzte mentale Ressourcen zur Interaktion mit einem SDS zur Verfügung stehen, lässt sich daher schlussfolgern, dass der Dialog möglichst effizient gestaltet sein sollte. Da der Kommunikationszweck häufig in der Erteilung eines klaren Auftrages an das SDS besteht und dieses als Werkzeug zur sicheren Interaktion verstanden wird, wird der Wunsch nach einem schnellen Erreichen des Bedienziels noch verstärkt. Verschiedene beobachtbare Verhaltensweisen der Nutzer in den durchgeführten Studien stützen die Annahme, dass sie selbst versuchen das Erreichen des Interaktionsziels zu beschleunigen. Beispielsweise können die Vereinfachungs- und Verdeutlichungsstrategien des Computer Talks (Kap. 4.1) oder die überlappenden Spracheingaben der Nutzer damit begründet werden, dass sie die Aufmerksamkeit und Zeit für die Sprachbedienung reduzieren wollten, um sich besser auf die Fahraufgabe konzentrieren zu können (vgl. Graham et al., 1999, S. 317-320). Auch das Sprachdialogsystem sollte seiner Verantwortung als Empfänger gerecht werden und den gemeinsamen (zeitlichen) Aufwand nur dann erhöhen, wenn es die Situation verlangt.

Verglichen zu den visuellen Feedbackanzeigen und dem systemseitigen Alignment, zeigte die dynamische Dialoggestaltung anhand der Erkennerkonfidenz den größten Effekt auf die Gebrauchstauglichkeit des SDS. Es kann geschlossen werden, dass von allen Manipulationen des Dialogverhaltens die Flexibilisierung des SGC nach dem LCE-Modell der vielversprechendste

Ansatz ist, da er den Dialog beschleunigt und somit die nutzerseitigen prozeduralen Interaktionsziele voll erfüllt.

Die sprachlichen Gesetzmäßigkeiten der zwischenmenschlichen Kommunikation scheinen auch bei der Interaktion mit maschinellen Gesprächspartnern Geltung zu besitzen, auch wenn sich ihre **prozeduralen Ziele** anders gewichten. Bezogen auf die Grice´schen Maximen, bedeutet dies, dass sich die Relevanz der einzelnen Richtlinien im Mensch-Maschine-Dialog während der Fahrt von alltäglichen zwischenmenschlichen Gesprächen unterscheidet. Während in alltäglichen Smalltalk-Konversationen der Effizienz keine bedeutende Rolle zukommt, konnte hier belegt werden, dass insbesondere die Maxime der Qualität und Art & Weise stark gewichtet vorliegen.

Weiterhin erscheint es sinnvoll, innerhalb des Dialogs Raum für Problembehebung zu schaffen (Clark & Brennan, 1991). Durch den Rückgriff auf die Collaborative Theory und die vorliegende Abhandlung soll Grice‘s Ansatz im Hinblick auf den Kommunikationsverlauf um kollaborative Empfängerhandlungen erweitert werden. Dem Empfänger kommt durch diese Rückmeldungen mehr Verantwortung zu, dem auch ein SDS im Dialog mit einem Menschen gerecht werden sollte. Meldet er Dialoginhalte zurück, so sollte er allerdings auch den Raum für effiziente Korrekturen geben.

Es kann geschlossen werden, dass der Flexibilisierung des SGC das größte Potential bei der nutzerorientierten Verbesserung der Usability von SDS zukommt. Insbesondere im Fahrtkontext, in der Sprachbedienung als Zweitaufgabe nicht die volle Aufmerksamkeit des Fahrers zuteilwerden kann, scheinen strategische oder soziale Aspekte der Kommunikation unwichtig verglichen zu der schnellen Erreichung des Bedienziels. In Situationen ohne mentale Beanspruchung und ohne direktes Bedienziel können dagegen auch eine intensivere Etablierung des Common Grounds und die Vernachlässigung der Effizienz in Mensch-Maschine-Interaktionen beobachtet werden. So sei auf die in Kapitel 4.1 erläuterten Konversationsinhalte mit dem Museumsavatar Max verwiesen, die eher einem Smalltalkgespräch zwischen Menschen glichen, als einem Problemlösedialog (Kopp et al., 2008). Auch der Sprachassistent Siri des iPhone 4GS[36] (Apple Inc., 2012) kann hier als Beispiel aufgeführt werden, der auch auf ungewöhnliche Anfragen (z. B. *Willst du mich heiraten?*) Antworten kennt. In Abgrenzung zu den spielerischen Dialogen mit diesen beiden Systemen, begünstigen Dialoge mit automotiven SDS, die zumeist mit nutzerseitiger

[36] Der Siri Sprachassistent des iPhone 4S nutzt die sprachliche Modalität, um Funktionen des Smartphones zu bedienen und stößt auf breites Interesse. Im Rahmen eines Dialoges können Nutzer z.B. Textnachrichten versenden oder Termininformationen natürlich-sprachlich abrufen.

mentaler Beanspruchung aufgrund der Fahrsituation einhergehen, das Monitoring and Adjustment Model (vgl. Horton & Keysar, 1996). Da die antizipierte Systemkompetenz nicht die Sprachproduktionsplanung beeinflusst, sollten Dialoge möglichst erwartungskonform gestaltet werden, um ein explizieren des CG zu vermeiden. Je nach Verfügbarkeit der Ressourcen scheinen beide Modelle zur Etablierung des CG innerhalb der Mensch-Maschine-Interaktion Anwendung zu finden.

Angeschlossene Regressionsanalysen belegten, dass die Zuverlässigkeit und die Angemessenheit der Dialoglänge die Zufriedenheit mit dem System vorhersagen konnten. Somit konnten sowohl ein Effektivitäts- als auch ein Effizienzparameter zur Aufklärung der Varianz beitragen. Es unterstützt die latente Annahme, dass sich eine hohe Nutzerzufriedenheit nur bei gegebener Systemeffektivität und -effizienz ergeben kann. Aufbauend auf diesen Erkenntnissen wird für automotive SDS eine **Zwei-Faktoren-Theorie der Usability** postuliert. In Analogie an Herzbergs Zwei-Faktoren-Theorie nach Ulich (2005) verdichtet sich die These, dass es sich bei der Effektivität und Effizienz um sogenannte Hygienefaktoren handelt, welche bei positiver Ausprägung die Entstehung von Unzufriedenheit verhindern, aber nicht zur Zufriedenheit beitragen können (siehe Abbildung 55Abbildung 55: Zwei-Faktoren-Theorie der Usability). Diese Faktoren werden bei Vorhandensein als selbstverständlich betrachtet, nur ihr Mangel hat schwere Konsequenzen auf die Zufriedenheit.

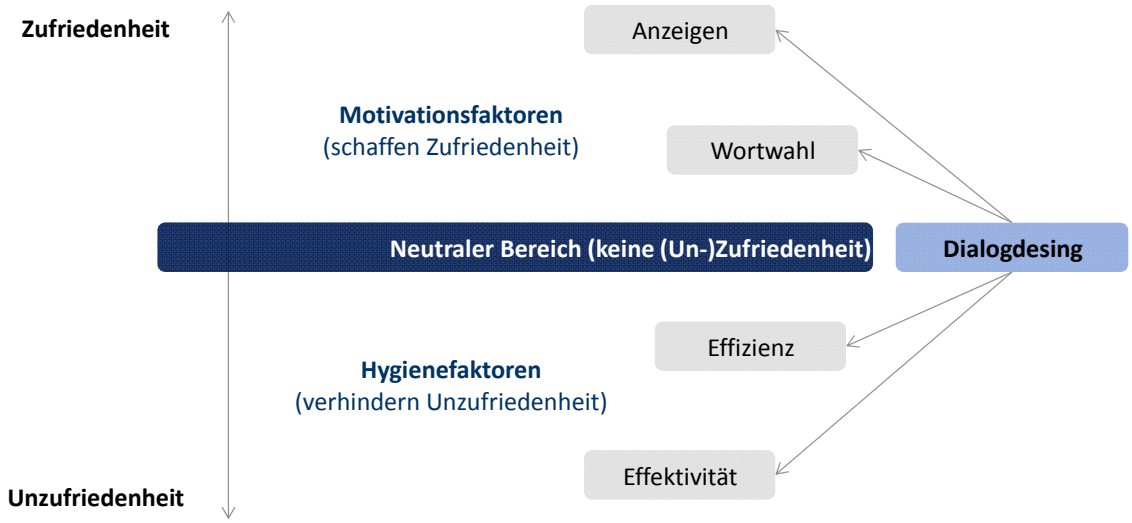

Abbildung 55: Zwei-Faktoren-Theorie der Usability

Insbesondere die geringere Effektstärke des Zufriedenheitskonstrukts in Studie II lässt sich über diesen Ansatz erklären. So zeigt das Dialogverhalten einen Einfluss auf die Effektivität und Effizienz mit beachtlicher Effektstärke. Da sich der Zufriedenheitseffekt nur durch die Vermittlung über die beiden Hygienefaktoren ergibt, zeigt dieser eine geringe praktische Relevanz. Eine Bezeich-

nung der Effizienz als Hygienefaktor soll jedoch ihren Einfluss auf die Gebrauchstauglichkeit eines SDS nicht schmälern. Die Effekte, die sich allein durch die Reduktion der Turn Takes in Studie II zeigten, sind vielversprechend.

Eine zentrale Erkenntnis der Arbeit ist dabei, dass auch das Dialogdesign dazu in der Lage ist, die Hygienefaktoren der Usability zu adressieren. Strategien aus dem zwischenmenschlichen Bereich, die allerdings primär die Zufriedenheit adressieren und den Dialog mit einem SDS vermenschlichen, können nur bei gegebener Effektivität und Effizienz des Systems erfolgreich sein. Insbesondere im automotiven Kontext mit den beschränkten Aufmerksamkeitskapazitäten scheinen den affektiven Dialogkomponenten geringere Bedeutungen zuzukommen.

Insgesamt konnte belegt werden, dass die Adaption bzw. **Individualisierbarkeit** bei der Gestaltung sprachlicher Interfaces über der Konsistenz stehen sollte. In Anlehnung an zwischenmenschliche Dialoge, die sich über die Dialog hinweg nur durch eine geringe Konsistenz auszeichnen, erscheint eine situations- und nutzerzentrierte Anpassung sinnvoller als eine statische Dialogführung oder Wortwahl. Die Befundlage legt die Schlussfolgerung nahe, dass in Abgrenzung zu grafischen Mensch-Maschine-Schnittstellen, die nicht-persistente Charakteristik von SDS dazu führt, dass Nutzer kein ausgeprägtes Konsistenzbild erstellen. So werden Dialoggestaltungsparameter, wie beispielsweise Formulierungsweisen, nicht in derselben Granularität memoriert wie grafische Bedienabläufe. Als Heuristiken zur Bewertung von SDS werden daher eher quantitative Parameter herangezogen (z. B. Zeit, Dialogschritte, Fehlerkennungen), die sich klar den Hygienefaktoren Effizienz und Effektivität zuordnen lassen.

Anhand der Erkenntnisse sollen im Folgenden Anstöße für zukünftige Untersuchungen abgeleitet werden.

7.2 Anstöße

Die Arbeit konnte weitere Optimierungsbereiche bestehender SDS identifizieren, die durch die Übertragungen zwischenmenschlicher Prinzipien der Collaborative Theory gelöst werden können. Weitere Ansätze sollen beispielhaft aufgezeigt werden, die das Grounding mit einem SDS erleichtern können und deren Umsetzung ein konsequenter nächster Schritt zur Gewährleistung von nutzerfreundlichen SDS darstellen würde.

Um dem interaktiven und kollaborativen Charakter der Kommunikation gerecht zu werden, sollten weitere Strukturen implementiert werden, die dem Nutzer ermöglichen das System zu unterbrechen oder überlappende Spracheingaben zu produzieren.

„Ein guter Alltagsdialog wickelt sich nie, niemals so ab wie auf dem Theater: mit Rede und Gegenrede. Das ist eine Erfindung der Literatur. Ein Dialog des Alltags kennt nur Sprechende – keinen Zuhörenden […]“ (Tucholsky, 1927/61, 713 f.)

Tucholsky (ebd.) verdeutlicht, dass eine einfache Gestaltung eines Dialoges durch begrenzte Zeiträume für die Nutzerein- oder Systemausgabe nicht zielführend ist und die menschliche Dialogführung nicht abbilden kann. Durch das sequentielle Dialogdesign ist der Nutzer bisher bei jeder Interaktion gezwungen zu warten bis eine Systemausgabe beendet und das System bereit für eine Spracheingabe ist. Verschiedenste zitierte und durchgeführte Untersuchungen können belegen, dass diese starre und unnatürliche Interaktionsstruktur Fehlbedienungen begünstigen und sich negativ auf die Effizienz des Dialoges auswirken. Insbesondere bei dem initialen Kommando konnte gezeigt werden, dass eine Statusvisualisierung und/oder eine akustische Tonfolge nicht ausreichen, um die zu frühe Spracheingabe des Nutzers zu verhindern. Die Implementation eines realen **Barge-Ins** würde es dem Nutzer erlauben die Effizienz des Dialoges selbstständig zu erhöhen und frei zu entscheiden, wann Ressourcen für die Interaktion zur Verfügung stehen. Diese Umsetzung bietet Potential, sowohl die Vorteile des Inhaltsfeedbacks als auch des Alignments im Grounding-Kontext zu betonen. Durch die Möglichkeit zur Unterbrechung könnten Nutzer den Zeitpunkt der Sprechbeiträge aktiv kontrollieren. Die Anzeige der Inhalte und die Angleichung würden als Verständnisevidenzen und damit der Beschleunigung des Dialogs dienen. Insbesondere bei automotiven SDS ließen sich durch diese erhöhte Steuerbarkeit des Dialogs positive Effekte auf die Gebrauchstauglichkeit erwarten. Für eine Weiterentwicklung der Systeme hin zu kollaborativen Empfängern ist die Implementation von umfassenden Barge-In-Strategien daher unabdingbar.

Den zweiten Optimierungsbereich stellt der nicht zufriedenstellende Umgang aktueller Sprachdialogsysteme mit **hyperartikulierten Äußerungen** dar. Er kann als ein weiteres Feld identifiziert werden, in dem eine zwischenmenschlich funktionale Strategie von Nutzern auf die Interaktion mit SDS übertragen wird und bisher Effektivitätseinbußen verursacht. Studie II konnte belegen, dass sich die überbetonten Eingaben nicht nur nach Fehlererkennungen, sondern auch dann beobachten lassen, wenn Nutzer mit dem Dialogablauf nicht zufrieden waren. Gerade in diesen Problemfällen sollte die Effektivität der Systeme gewährleistet sein, um einem Teufelskreis der Fehlererkennung und nutzerseitiger Frustration zu entgehen. Zukünftige Spracherkenner sollten demzufolge nicht nur mit flüssig eingesprochener Sprache trainiert werden, sondern auch hyperartikulierte Äußerungen enthalten. Um der Überbetonung entgegenzuwirken, könnten Nutzereingaben durch die Systemausgaben so beeinflusst werden, dass der Spracherkenner sie am besten versteht. Als Beispiel kann das Sprachdialogsystem des Chevrolet Volts (2010) gelten, das durch ein Vorsprechen der Kommandos nicht nur die hyperartikulierten Eingaben reduziert, son-

dern damit auch die Systemkompetenzen transparent vermittelt. Insgesamt sollte jedoch eine gute Dialoggestaltung den frustrationsbedingten Hyperartikulationen entgegenwirken.

Weiterhin könnten die Gesprächsinitiierungskosten mit einem SDS reduziert werden, indem der **PTT-Knopf** von einer Schlüsselwortaktivierung abgelöst würde. Insgesamt konnte in allen Nutzerstudien beobachtet werden, dass die multifunktionale Belegung des PTT-Knopfes häufig missverstanden wurde. So konnten einerseits Probanden beobachtet werden, die vor jedem Dialogschritt erneut die Taste drückten, auch wenn das System bereits aktiv war. Andererseits wurde die erneute Aktivierung nach einem bereits abgeschlossenen Dialog häufig vergessen und Probanden sprachen Kommandos ein, ohne zuvor den PTT-Knopf betätigt zu haben. Die bisherige doppelte Belegung des Knopfes zur Aktivierung und Unterbrechung des Dialogs kann hinsichtlich der Intuitivität als nicht zufriedenstellend bewertet werden.

Insgesamt erscheint der Ansatz, etablierte Prinzipien der zwischenmenschlichen Kommunikation auf SDS zu übertragen und diese damit zu kollaborativen Gesprächspartnern zu machen, als vielversprechend. Eine spezielle **Sprache** mit SDSs zu entwickeln (Tomko & Rosenfeld, 2006, Schiel, 2006) wiederspricht der Grundidee von Sprache als intuitive Bedienmodalität. In einer Welt, in der Wörter bereits eine Bedeutung haben, erscheinen Aufforderungen wie „*Für Braunschweig sagen Sie Zeile 2.*“ grotesk (Balentine et al.,1999). Schiel (2006) betrachtet bereits die herstellerübergreifende Implementation von globalen Kommandos wie *Hilfe* oder *Abbrechen* als einen ersten Schritt in Richtung standardisierte Sprachentwicklung. Allerdings finden jene Kommandos nutzerseitig kaum Verwendung, da sie keinem zwischenmenschlichen Konzept entsprechen. So konnte beobachtet werden, dass das integrierte Hilfe-Kommando in Dialogabläufen kaum genutzt wird (siehe auch Mann, 2010). Um natürliche Dialoge zu gestalten, sollten Systeme dem Nutzer vielmehr ermöglichen auch negative Evidenzen zu geben (Brennan, 1998). Durch das Zulassen von Nachfragen (*Wie bitte?*) könnte es dem Nutzer nicht nur gestattet werden nach zusätzlichen Informationen oder Erklärungen zu fragen, sondern auch einen leichteren Wieder-Einstieg in den Dialog nach einem komplexen Fahrmanöver zu finden.

Betrachtet man das hohe Ausmaß der nutzerseitigen **Anpassung** an das Systemvokabular, so muss stets gewährleistet sein, dass ein SDS sein eigenes Vokabular versteht (Brennan, 1998). Verwendet ein Sprecher einen bestimmten Begriff, so kann der Gesprächspartner stets davon ausgehen, dass er diesen auch versteht. Alles andere stellt einen groben Verstoß gegenüber zwischenmenschlichen Kommunikationsprinzipien der Collaborative Theory dar.

Auch in absehbarer Zukunft ist nicht mit der perfekten maschinellen Spracherkennung zu rechnen. Auch wenn der menschliche Dialog im Rahmen dieser Arbeit als das **Vorbild** für die Gestaltung eines SDS postuliert wurde, so kann dieser auch nicht als fehlerfrei gelten. Auch innerhalb von zwischenmenschlichen Gesprächen lassen sich Missverständnisse beobachten.

"Of course, understanding can never be perfect." (Clark & Brennan, 1991, S. 129).

Dennoch treten Erkennungsfehler, wie in Kapitel 2.2.3 beschrieben, deutlich weniger häufig auf und gefährden nur selten die Erreichung des Kommunikationsziels. Auch bei ungünstigen Bedingungen können menschliche Dialogpartner zumeist erfolgreich kommunizieren, da die Auflösung von Missverständnissen und eine inhärente Fehlerkorrektur ein Teil der Gesprächsführung sind (Hamerich, 2009). Verglichen mit dem Mensch-Maschine-Dialog, wissen Gesprächspartner häufig, warum eine Äußerung nicht verstanden wurde und leiten entsprechende Reparaturmechanismen ein. Bei aktuellen SDS sind jegliche unerwartete Systemreaktionen für den Nutzer jedoch nicht transparent. Es ist für ihn beispielsweise nicht ersichtlich, ob ein Kommando nicht verstanden wurde, da es nicht im Systemwortschatz vorhanden ist oder die Konfidenz zur korrekten Erkennung nicht ausreichte. Die Unterstützung von **Dialogreparaturstrategien**, sollte gerade bei Erkennungsproblemen ausgebaut werden. Der folgende Abschnitt soll einen Ausblick auf zukünftige Interkationen von Nutzern und SDS geben.

7.3 Ausblick

Im Moment kann aufgrund der noch geringen **Verbreitung und Akzeptanz** von automotiven Sprachdialogsystemen davon ausgegangen werden, dass Menschen sich bei dem Empfängermodell, das sie von ihrem Dialogpartner erstellen, sehr unsicher sind. Sie haben ein geringe Vorstellung davon, was das SDS leisten oder verstehen kann und erstellen aus diesem Grund ein sehr konservatives initiales Modell des maschinellen Gesprächspartners. In den durchgeführten Studien konnte mehrfach beobachtet werden, dass sich Nutzer über die technologische Funktionsweise maschineller Spracherkennung nicht im Klaren sind. So wurden beispielsweise Substitutionsfehler nutzerseitig häufig als Vorschläge des Systems missinterpretiert. Die meisten Nutzer sehen diese Art von Fehler also nicht als ein Erkennungsproblem, sondern glauben, dass das System ihnen eine Alternative vorschlagen möchte, worauf sie in den meisten Fällen mit Reaktanz reagieren.

Für zukünftige SDS kann jedoch bei der gegenwärtigen Verbreitung davon ausgegangen werden, dass sich Nutzererwartungen und Systemkompetenzen immer mehr angleichen werden. Eine große Rolle spielt dabei neben den erkennungstechnologischen und gestalterischen Fortschritten der SDS, dass Nutzer immer mehr Erfahrung mit Sprachdialogsystemen sammeln und deren Kompetenzen und Grenzen somit realistischer einschätzen können. Bereits in Studie III konnte belegt werden, dass Experten die Qualität von SDS realistischer beurteilen können. Je mehr Erfahrungen Nutzer im Bereich der mobilen Endgeräte mit SDS sammeln können, umso größer wird nicht nur deren Fähigkeit im Umgang mit dieser Technologie, sondern umso realistischer erfolgt auch deren Aufbau eines Empfängermodells. Der Siri Sprachassistent (Apple Inc., 2012) wird maßgeblich dazu beitragen, dass eine breite Masse an Endkunden bereits Zugang zu ma-

schineller Sprachtechnologie erhält. Aufgrund dieser aktuellen Entwicklungen wird sich die Kluft zwischen den nutzerseitigen Erwartungen und Erfahrungen auch bei automotiven Sprachdialogsystemen verringern. In weniger als einer Dekade werden Nutzer ein gefestigtes Sprachregister für SDS ausgebildet haben, das die derzeitige Definition des Computer Talks spezifiziert.

Die Abhandlung konnte insgesamt belegen, dass der Einstieg in die Sprachbedienung zum jetzigen Zeitpunkt durch die Anwendung zwischenmenschlicher Kommunikationsprinzipien erleichtert werden kann. Eine transparente Vermittlung der Systemkompetenzen und eine kollaborative Gesprächsführung half die anfängliche Unsicherheit im Gespräch mit einer Maschine zu reduzieren. Da Menschen funktionale Strategien der zwischenmenschlichen Kommunikation ohnehin mit in den Mensch-Maschine-Dialog bringen, muss eine Anwendbarkeit der selbigen gewährleistet sein. Ebenso erwarten sie den Gebrauch jener Dialogstrategien auch von ihrem maschinellen Gesprächspartner.

Für das **Design** von automotiven Sprachdialogsystemen sollten somit die Maximen von Grice (1975), ergänzt durch die kollaborative Theorie und die hier erarbeiteten Richtlinien, Anwendung finden. Dabei sollte die in Kap. 7.1 empfohlene Gewichtung im Fahrtkontext nicht außer Acht gelassen werden. Denn die mental beanspruchende Fahrtsituation und die Sicht des SDS als Werkzeug führen dazu, dass strategischen Aspekten verglichen zu effizienzsteigernden Strategien der Dialogführung nur eine untergeordnete Rolle zukommt. Der Transfer zwischenmenschlicher Strategien sollte somit nicht dazu verwendet werden, den Mensch-Maschine-Dialog affektiv aufzuwerten oder zu vermenschlichen. Es sollten vielmehr Prinzipien übernommen werden, die Menschen anwenden, um Interaktionsziele möglichst effizient zu erreichen. Insbesondere die Collaborative Theory, die primär auf der deskriptiven Beschreibung zwischenmenschlicher Dialoge in Problemlöseszenarien fußt, bot sich für diesen Transfer an, um eine Steigerung der Effizienz und Effektivität trotz der begrenzten Systemkompetenzen zu realisieren.

So konnte bereits in der vorliegenden Arbeit mit adäquatem technischem Aufwand ein menschlicherer Dialog erreicht werden. Dies bedeutet nicht zwingend, dass SDS als Gesprächspartner auf Augenhöhe empfunden werden. Das hierarchische Gefälle, welches zwischen dem menschlichen Nutzer und dem System besteht, bleibt über jeden Dialog hinweg präsent. So besitzen Menschen eine vielfach höhere Sprach- und Anpassungskompetenz, die sie auch bereit sind in der Mensch-Maschine-Kommunikation anzuwenden, um die Dialog effizienter oder effektiver zu gestalten. Durch kollaborative Empfängerhandlungen sollte aber auch das SDS zu einer gemeinsamen Aufwandsreduktion beitragen. Ein Sprachagent, der zügig und zuverlässig die Wünsche des Autofahrers erfüllt und dabei sein Verhalten an die Stellschrauben der zwischenmenschlichen Kommunikation anlehnt, könnte die Vorteile der Sprachbedienung im automotiven Kontext voll zur Geltung bringen und Dialoge ermöglichen, wie sie bereits 1982 mit KITT gezeigt wurden (Fenady et al., 1982).

8 Literaturverzeichnis

Ang, J., Dhillon, R., Krupski, A., Shriberg, E. & Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proceedings of ICSLP-2002*, (S. 2037-2040). Abgerufen von http://www.icsi.berkeley.edu/ftp/global/pub/speech/papers/icslp2002-emotion.pdf

Apple Inc. (2012). *Siri. dein Wunsch ist ihm Befehl.* Abgerufen von http://www.apple.com/de/iphone/features/siri.html?cid=mc-iphone-de-g-rii-siri&sissr=1

Baber, C. & Strammers, R. (1989). Is it natural to talk to Computers? An Experiment using the "Wizard of Oz" Technique. In E. D. Megaw (Hrsg.) *Contemporary Ergonomics* (S. 234-239). Taylor & Francis, London.

Bailenson, J. & Yee, N. (2005). Digital chameleons-automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16*(10), 814-819.

Balentine, B., Morgan, D. P. & Meisel, W. S. (2001). *How to Build a Speech Recognition Application: Second Edition: A Style Guide for Telephony Dialogues.* SanRamon CA: Enterprise Integration Group.

Bangor, A., Kortum, P. & Miller, J. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction, 24*(6), 574-594.

Bavelas, J., Black, A., Lemery, C. & Mullett, J. (1986). "I show how you feel": motor mimicry as a communicative act. *Journal of Personality and Social Psychology, 50*(2), 322-329.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*(2), 145-204.

Bell, L., Gustafson, J. & Heldner, M. (2003). Prosodic adaptation in human-computer interaction. In *Proceedings of the International Congress of Phonetic Sciences,* (S. 2453-2456). Abgerufen von http://www.speech.kth.se/ctt/publications/papers03/icphs03_2453.pdf

Bengler, K. (1995). *Gestaltung und experimentelle Untersuchung unterschiedlicher Präsentationsformen von Wegleitungsinformationen in Kraftfahrzeugen.* Regensburg: S. Roder Verlag.

Bengler, K. (2001). Aspekte der multimodalen Bedienung und Anzeige im Automobil. In T. Jürgensohn & K.-P. Timpe (Hrsg.), *Kraftfahrzeugführung* (S. 195-205). Berlin: Springer Verlag.

Bernsen, N. O. & Dybkjær, L. (2001). Usability evaluation in spoken language dialogue systems. In *Proceedings of the ACL workshop on Evaluation Methodologies for Language and Dialogue Systems*, (S. 9-18). Abgerufen von http://delivery.acm.org/10.1145/1120000/1118055/p2-dybkjaer.pdf?ip=194.114.62.72&acc=OPEN&CFID=117453951&CFTOKEN=49341234&__acm__=1348130502_00068c94cdd62a5ddb54d4e92bba9785

Bernsen, N. O. & Dybkjær, L. (2004). Building usable spoken dialogue systems. Some approaches. *International Journal for Language Data Processing* 28(2), 111-131.

Blattner, M. M., Sumikawa, D. A. & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human–Computer Interaction, 1*, 11-44.

Bortfield, H. & Brennan, S. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes, 23*(2), 119-147.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Ausg.). Heidelberg: Springer Medizin Verlag.

Boyce, S. J. (2008). User interface design for natural language systems: From research to reality. In D. Gardener-Bonneau & H. Blanchard (Hrsg.), *Human factors and voice interactive systems* (2. Ausg., S. 43-80). Berlin: Springer.

Bradac, J., Mulac, A. & House, A. (1988). Lexical diversity and magnitude of convergent versus divergent style shifting perceptual and evaluative consequences. *Language and Communication, 8*(3-4), 213-228.

Bradley, D. R. & Russell, R. L. (1998). Some cautions regarding statistical power in split-plot designs. *Behavior Research Methods, Instruments & Computers, 30*(3), 462-477.

Bradlow, A. R. & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112*, 272–284.

Brandstätter, E. (1999). Konfidenzintervalle als Alternative zu Signifikanztests. *Methods of Psychological Research Online, 2*, 1-17.

Branigan, H., Pickering, M. & Cleland, A. (2000). Syntactic coordination in dialogue. *Cognition, 25*(2), 13-25.

Branigan, H., Pickering, M., Pearson, J. & McLean, J. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics, 42*(9), 2355-2368.

Branigan, H., Pickering, M., Pearson, J., McLean, J. & Nass, C. (2003). Syntactic alignment between computers and people: the role of belief about mental states. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*, (S. 186-191). Abgerufen von http://csjarchive.cogsci.rpi.edu/proceedings/2003/pdfs/55.pdf

Bratzel, S. (2011). *"i-Car": Die junge Generation und das vernetzte Auto. Eine empirische Studie zu den Einstellungen und Verhaltensmustern der 18-25 Jährigen in Deutschland.* Bergisch Gladbach: Center of Automotive Management.

Brennan, S. E. (1991). Conversations with and through computers. *User Modeling and User-Adapted Interaction, 1*(1), 67-86.

Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. In *Proceedings of the International Symposium on Spoken Dialogue*, (S. 41-44). Abgerufen von http://www1.cs.columbia.edu/~julia/papers/brennan96.pdf

Brennan, S. E. (1998). The grounding problem in conversation with and through computers. In R. J. S. R. Fussell (Hrsg.), *Social and cognitive psychological approaches to interpersonal communication* (S. 201-225). Hillsdale, NJ: Lawrence Erlbaum.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. T. Tanenhaus (Hrsg.), *World situated language use: psycholinguistic, linguistic and computational perspectives on bridging the product and action traditions* (S. 95–129). Cambridge: MIT Press.

Brennan, S. E. & Hulteen, E. (1995). Interaction and feedback in a spoken language system: A theoretical framework, *Knowledge-Based Systems*, *8*(2), 143-151.

Brennan, S. E. & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398.

Brennan, S. & Clark, H. (1996). Conecptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(6), 1482-1493.

Brewster, S. A. (2002). Non-speech auditory output. In J. A. Jacko & A. Sears (Hrsg.), *Human-Computer Interaction Handbook* (S. 220-239). Mahwah, NJ: Lawrence Erlbaum Associates.

Brey, T. & Salmen, A. (2003). Multimodales Interaktionsmanagement. *Konzeptstudie für die Robert Bosch GmbH*.

Brooke, J. (1996). SUS - A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & A. L. McClelland (Hrsg.), *Usability Evaluation in Industry* (S. 189-194). London: Taylor and Francis.

Brumby, D. P., Davies, S. C., Janssen, C. P. & Grace, J. J. (2011). Fast or safe? How Performance Objectives Determine Modality Output Choices while Interacting on the Move. In *Proceedings of the CHI 2011*, (S. 473-482). Abgerufen von http://www.uclic.ucl.ac.uk/people/d.brumby/publications/Brumby.2011.CHI.pdf

Bubb, H. (1993). Systemergonomische Gestaltung. In H. Schmittke (Hrsg.), *Ergonomie* (S. 391-420). München: Hanser.

Bubb, H. (2003). Fahrerassistenz primär ein Beitrag zum Komfort oder für die Sicherheit? *VDI-Bericht Nr. 1768*, (S. 25-44). Düsseldorf: VDI-Verlag.

Burnett, G. & Joyner, S. (1997). An assessment of moving map and symbol-based route guidance systems. In Y. Ian Noy (Hrsg.), *Ergonomics and safety of intelligent driver interfaces* (S. 115-136). Mahwah: Lawrence Erlbaum Associates.

Caird, J. K., Willness, C. R., Steel, P. & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention, 4*, 1282-1293.

Cameron, H. (2000). Speech at the interface. In *Proceedings of the COST249 Workshop: "Voice Operated Telecom Services - do they have a bright future?",* (S. 1-7). Abgerufen von http://kom.aau.dk/~tb/cameron.pdf

Cathcart, J., Carletta , J. & Klein, E. (2003). A shallow model of backchannel continuers in spoken dialogue. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 1*, (S. 51-58). Abgerufen von http://delivery.acm.org/10.1145/1070000/1067816/p51-cathcart.pdf?ip=194.114.62.72&acc=OPEN&CFID=117453951&CFTOKEN=49341234&__acm__=1348132554_4ba57056ce18d85c3bcba57726945b84

Chartrand, T. & Bargh, J. (1999). The cameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910.

Chu-Carroll, J. & Nickerson, J. S. (2000). Evaluating automatic dialogue strategy adaptation for a spoken dialogue system. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (S. 202-209). Seattle, Washington: Morgan Kaufmann Publishers Inc.

Clark, H. H. (1996). *Using Language.* Cambridge, MA: Cambridge University Press.

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Hrsg.), *Perspectives on socially shared cognition* (S. 127-149), Washington, DC: APA Books.

Clark, H. H. & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39.

Clark, H. & Krych, M. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62-81.

Cohen, M., Giangola, J. & Balogh, J. (2004). *Voice user interface design.* Boston: Addison-Wesley.

Core, M. G. & Moore, J. D. (2004). Robustness versus fidelity in natural language understanding. In: R. Porzel (Hrsg.), *HLT-NAACL 2004 Workshop: 2nd Workshop on Scalable Natural Language Understanding* (S. 1–8). Boston: Association for Computational Linguistics.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Curin, J., Labsky, M., Macek, T., Kleindienst, J., Young, H., Thyme-Gobbel, A., … König, L. (2011). Dictating and Editing Short Texts while Driving: Distraction and Task Completion. In *Proceedings of the AutomotiveUI 2011 - 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* (S. 13-20). Abgerufen von http://www.auto-ui.org/11/docs/AUI2011_proceedings.pdf

Cutler, A. & Butterfield, S. (1990). Durational cues to word boundaries in clear speech. *Speech Communication, 9*(5-6), 485-495.

Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993). Wizard of Oz studies-why and how. *Knowledge-Based Systems, 6*(4), 258-266.

Deutsches Institut für Normung. (1997). *Sicherheit von Maschinen - Ergonomische Anforderungen an die Gestaltung von Anzeigen und Stellteilen - Teil 2: Anzeigen; Deutsche Fassung EN 894-2.* Berlin: Beuth.

Deutsches Institut für Normung. (1998). *Europäische Norm DIN EN ISO 9241-11. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11: Anforderungen an die Gebrauchstauglichkeit.* Berlin: Beuth.

Deutsches Institut für Normung. (2006). *Europäische Norm DIN EN ISO 9241-110. Ergonomie der Mensch-System-Interaktion. Teil 110: Grundsätze der Dialoggestaltung.* Berlin: Beuth.

Dix, A., Finlay, J., Abowd, G. D. & Beale, R. (2004). *Human-Computer Interaction* (3. Ausg.). Upper Saddle River, NJ: Prentice Hal.

Düker, H. (1963). Über reaktive Anspannungssteigerung. *Zeitschrift für experimentelle und angewandte Psychologie, 10*, 46-72.

Engelhardt, D. (2009). Systemkriterien für zukünftige Sprachdialogsysteme im Automobil. *Unveröffentlichte Diplomarbeit Volkswagen AG*.

Ergoneers. (2011). *Dikablis & D-Lab*. Abgerufen von http://www.ergoneers.com/de/products/dlab-dikablis/overview.html

Everitt, B. S. (1996). *Making Sense of Statistics in Psychology.* Oxford: Oxford University Press.

Fenady, G., Kolbe, W., Hayers, S., Kowalski, B. L. & Laidman, H. S. (Regisseure). (1982). *Knight Rider* [Kinofilm]. New York: NBC Universal.

Ferguson, C. (1975). Toward a characterization of English foreigner talk. *Anthropological Linguistics, 17*(1), 1-14.

Ferguson, C. A. (1977). Baby talk as simplified register. In C. E. Snow, C. A. Ferguson (Hrsg.), *Talking to Children: Language Input and Acquisition (S. 219-236).* Cambridge University Press.

Fillmore, C. (1981). Pragmatics and the Description of Discourse. In P. Cole (Hrsg.), *Radical Pragmatics* (S. 143-166). New York: Academic Press.

Fischer, K. (1999). Repeats, reformulations, and emotional speech: Evidence for the design of human-computer speech interfaces. In H.-J. Bullinger & J. Ziegler (Hrsg.), *Human-Computer Interaction: Ergonomics and User interfaces, Proceedings of the 8th International Conference on Human-Computer Interaction*, *Vol. 1*, S. 560-565. Lawrence Erlbaum Ass, London.

Fischer, K. (2006). *What Computers Talk is and isn´t, Human-Computer Conversation as Intercultural Communication* (Bd. 17). Saarbrücken: AQ-Verlag.

Fowler, C., Brown, J., Sabadini, L. & Weihing, J. (2003). Rapid access to speed gestures in perception: ecidence from choice and simple response time task. *Journal of Memory and Language, 49*(3), 396-413.

Fraser, N. & Gilbert, G. (1991). Simulating speech systems. *Computer Speech and Language, 5*, 81-99.

Frost & Sullivan. (01. Februar 2012). Strategic Analysis of European and North American Automotive Human Machine Interface Market - Voice Control Interface, Steering Wheel Controls, and Multifunctional Knob. Abgerufen von http://www.frost.com/prod/servlet/report-toc.pag?ctxixpLink=FcmCtx3&searchQuery=Basic+Voice+Interface&repid=M6CF-01-00-00-00&bdata=aHR0cDovL3d3dy5mcm9zdC5jb20vc3JjaC9jYXRhbG9nLXNlYXJjaC5kbz9xdWVyeVRleHQ9QmFzaWMrVm9pY2UrSW50ZXJmYWNlQH5AU2VhcmNoIFJlc3VsdHNA.

Furnas, G. W., Landauer, T. K., Gomez, L. M. & Dumais, S. T. (1987). The vocabulary problem in human-system. *Commun. ACM 30, 11*, 964-971.

Garrod, S. & Clark, H. H. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes, 8*(1), 101-126.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction, 2*(2), 167-177.

Gergle, D., Kraut, R. & Fussell, S. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology, 23*(4), 491-517.

Gieselmann, P. & Stenneken, P. (2006). How To Talk to Robots: Evidence from User Studies on Human-Robot Communication. In *Proceedings of the Workshop Hansewissenschaftskolleg,* (S. 68-78). Abgerufen von http://pub.uni-bielefeld.de/download/1989997/2487171

Giles, H. & Powesland, P. (1975). *Speech Style and Social Evaluation.* London: Academic Press.

Giles, H., Coupland, N. & Coupland, J. (1991). Accomodation theory: communication, context, and consequence. In H. Giles, J. Coupland & N. Coupland (Hrsg.), *Contexts of Accomodation: Developments in Applied Sociolinguistics* (S. 1-68). Cambridge: Cambridge University Press.

Goldstein, E. B. (1997). *Wahrnehmungspsychologie. Eine Einführung.* Heidelberg: Spektrum Akademischer Verlag.

Goleman, D. (2006). *Social Intelligence: the new science of human relationships.* New York: Bantam Books.

Gorrell, G. (2003). *Language Modelling for Spoken Dialogue Systems; Grammar-Based and Robust Approaches Compared and Contrasted.* Abgerufen von http://www.speech.kth.se/~matsb/speech_rec_course_2003/papers/Genevieve_G/speech_course_essay.pdf

Goulati, A. & Szostak, D. (2011). User experience in speech recognition of navigation devices: an assessment. In *Proceedings of the 13th International Conference an Human Computer Interaction with Mobile Device and Services*, S. 517-520. New York: ACM.

Graham, R., Aldrigde, L., Carter, C. & Lansdown, T. (1999). The design of incar speech recognition interfaces for usability and user acceptance. In D. Harris (Hrsg.), *Engineering psychology and cognitive ergonomics: Job design, product design and human-computer interaction* (S. 313-320). Aldershot: Ashgate.

Green, D. M. & Swets, J. (1966). *Signal Detection Theory and Psychophysics.* New York: Wiley.

Green, P. (2001). Variations in Task Performance Between Younger and Older Drivers. In *Proceedings of the Association for the Advancement of Automotive Medicine Conference on Aging and Driving,* (S. 1-12) Abgerufen von http://www.umich.edu/~driving/publications/AAAMC2001.pdf

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Hrsg.), *Syntax and Semantics III: Speech Acts* (S. 243-265). New York: Academic Press.

Gries, S. (2005). Syntactic priming: a corpus-based approach. *Journal of Psycholinguistics Research, 34*(4), 365-399.

Grimm, G. (1812/1815). *Kinder- und Hausmärchen.* Berlin.

Halliday, M. & Hasan, R. (1989). *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective.* Oxford: Oxford University Press.

Hamerich, S. W. (2009). *Sprachbedienung im Automobil: Teilautomatisierte Entwicklung benutzerfreundlicher Dialogsysteme.* Berlin Heidelberg: Springer.

Hanrieder. (2004). Sprachbedienung im KFZ- Eine Erfolgsgeschichte. In *Proceedings of the GI Jahrestagung,* (S. 220-224). Abgerufen von http://subs.emis.de/LNI/Proceedings/Proceedings50/GI-Proceedings.50-46.pdf

Hanrieder, G. & Hamerich, S. W. (2004). Modeling generic dialog applications for embedded systems. In *Proceedings of the INTERSPEECH,* (S. 237-240). Abgerufen von http://www.mirlab.org/conference_papers/International_Conference/ICSLP%202004/contents/TuA_pdf/TuA1302p/TuA1302p.18_p427.pdf

Hänsler, E. & Schmidt, G. (2004). *Acoustic Echo and Noise Control.* New York: John Wiley & Sons.

Harbluk, J., Burns, P., Lochner, M. & Trbovich, P. (2007). Using the lane-change test (LCT) to assess distraction: Tests of visual-manual and speech-based operation of navigation system interfaces. In *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design,* (S. 16-22). Abgerufen von http://drivingassessment.uiowa.edu/DA2007/PDF/004_Harbluk.pdf

Harris, R. A. (2005). *Voice Interaction Design: Crafting the New Conversational Speech Systems.* San Francisco: Morgan Kaufman.

Hart, S. G. & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Hrsg.), *Human mental workload* (S. 139-183). Amsterdam: Elsevier.

Hassel, A. L. (2006). *Adaption und Evaluation von Sprachbediensystemen im Automobilbereich.* Berlin: Logos Verlag.

Haywood, S., Pickering, M. & Branigan, H. (2005). Do speakers aviod ambiguities during dialogue? *Psychological Science, 16*(5), 362-366.

Hedicke, V. (2002). Multimodalität in Mensch-Maschine-Schnittstellen. In K.-P. Timpe, T. Jürgensohn & H. Kolrep-Rometsch (Hrsg.), *Mensch-Maschine-Systemtechnik* (S. 203-232). Düsseldorf: Symposion Publishing.

Heller, O. (1985). Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). *Psychologische Beiträge, 27*, 509-519.

Herczeg, M. (1994). *Software-Ergonomie. Grundlagen der Mensch-Computer-Kommunikation.* Bonn: Addison-Wesley.

Hewitt, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., … Verplank, W. (1992). *ACM SIGCHI Curricula for Human-Computer Interaction*. Abgerufen von http://old.sigchi.org/cdg/

Hipp, C., Paulke, S., Peissner, M. & Steimel, B. (2008). *Kochbuch für gute Sprachapplikationen.* Stuttgart: Fraunhofer IRB Verlag.

Hirasawa, J., Nakano, M., Kawabata, T. & Aikawa, K. (1999). Effects of system barge-in responses on user impressions. In *Proc. EUROSPEECH'99* (S. 1391-1394). Budapest, Ungarn.

Hockey, B. A., Lemon, O., Campana, E., Hiatt, L., Aist, G., Hieronymus, J. & Gruenstein, A. (2003). Targeted Help for Spoken Dialogue Systems: Intelligent Feedback Improves Naive User's Performance. In *Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics,* (S. 147–154). Abgerufen von http://wayback.archive-it.org/1792/20100511213749/http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20030107286_2003133805.pdf

Hof, A. (2007). *Entwicklung eines adaptiven Hilfesystems für multimodale Anzeige-Bedienkonzepte im Fahrzeug.* Dissertation, Regensburg: Universität Regensburg.

Horton, W. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91-117.

Huber, R., Noth, E., Baliner, A., Buckow, J., Warnke, V. & Niemann, H. (1998). You beeb machine - emotion in automatic speech understanding systems. In *Proceedings of the Workshop on Text, Speech, and Dialog,* (S. 223-228). Abgerufen von reference.kfupm.edu.sa/content/y/o/you_beep_machine_____emotion_in_automati_100429.pdf

IBM. (2001). *IBM Websphere Voice Server Software Developers Kit (SDK) Programmes Guide* (2. Ausg.). Abgerufen von www.ims.uni-stuttgart.de/lehre/teaching/2001-SS/Dialogsysteme/ibm_voiceserver_progguide.pdf

International Organization for Standardization (2010). *ISO 26022:2010. Road vehicles – ergonomic aspects of transport informationand control systems – simulated lane change test to assess in-vehicle secondary task demand.* Abgerufen von http://isotc.iso.org/livelink/livelink?func=ll&objId=11560806&objAction=browse&viewType=1

Jelinek, F. (1998). *Statistical Methods for Speech Recognition.* Cambridge: MIT Press.

Johnstone, A., Berry, U., Nguyen, T. & Asper, A. (1995). There was a Long Pause: influencing turn-taking behaviour in human-human and human-computer dialogs. *International Journal Of Human-Computer Studies*, *44*, 383-411.

Jones, B. & Nachtsheim, C. J. (2009). Split-plot designs: what, why, and how. *Journal of Quality Technology, 41*(4), 340-361.

Junqua, J. C. (1993). The Lombard Reflex and its Role on Human Listeners and Automatic Speech Recognizers. *Journal of the Acoustic Society of America 93*, 510–524.

Kahneman, D. (1973). *Attention and Effort.* New York: Prentice Hall.

Kainer, M. (2007). *Entwicklung eines grafischen Hilfe- und Feedbackkonzepts für multimodale Bedienelemente im Fahrzeug.* Diplomarbeit, Graz: FH Joanneum.

Kaiser, S. (2009). *Kognitive und visuelle Ablenkung auf das Fahrverhalten.* Diplomarbeit, Braunschweig: TU Braunschweig.

Kass, R. & Finin, T. (1988). Modeling the User in Natural Language Systems. *Computational Linguistics, 14*(3), 5-22.

Kirakowski, J. (1994). *The use of questionnaire methods for usability assessment*. Abgerufen am 05. 01 2012 von http://sumi.ucc.ie/sumipapp.html

Kirakowski, J. & Corbett, M. (1993). SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, *24*, 210-212.

Knappe, G., Keinath, A. & Meinecke, C. (2006). Empfehlungen für die Bestimmung der Spurhaltegüte im Kontext der Fahrsimulation. *MMI-Interaktiv, 11*, 3-13.

Kopp, S., Allwood, J., Grammer, K., Ahlsen, E. & Stocksmeier, T. (2008). Modeling embodied feedback in virtual humans. In I. Wachsmuth & G. Knoblich (Hrsg.), *Modeling Communication With Robots and Virtual Humans* (S. 18-37). Berlin: Springer.

Kraiss, K. F. (2006). *Advanced Man-Machine Interaction: Fundamentals and Implementation* (Bd. 1). Berlin, Heidelberg: Springer-Verlag.

Krallmann, D. & Ziemann, A. (2001). Die Informationstheorie von Claude E. Shannon. In Krallmann, D. & Ziemann, A (Hrsg.), *Grundkurs der Kommunikationswissenschaft: mit einem Hypertext-Vertiefungsprogramm im Internet* (S. 21-34). München: Fink.

Krause, J. (1992). Natürlichsprachliche Mensch-Computer-Interaktion als technisierte Kommunikation: Die computer talk-Hypothese. In J. Krause & L. Hitzenberger (Hrsg.), *Sprache und Computer, Computer Talk* (S. 1-29). Hildesheim: Olms.

Krause, J. & Hitzenberger, L. (1992). *Computer Talk.* Hildesheim: Olms.

Kun, A., Paek, T. & Medenica, Z. (2007). The Effect of Speech Interface Accuracy on Driving Performance. In *Proceedings of the European Conference on Speech Communication and Technology*, (S. 1326-1329). Abgerufen von http://research.microsoft.com/en-us/um/people/timpaek/papers/driv_sim_interspeech07.pdf

Lee, J. D., Caven, B., Haake, S. & Brown, T. L. (2001). Speech-Based Interaction with In-Vehicle Computers: The Effect of Speech-Based E-Mail on Drivers' Attention to the Roadway. *Human Factors, 43*, 631-640.

Leiber, P. (2010). *Ergonomische Produktgestaltung am Beispiel mobiler Geräte im interkulturellen Vergleich: China-Deutschland-USA.* Dissertation, Technische Universität Chemnitz: Wissenschaftliche Schriftenreihe des Instituts für Betriebswissenschaften und Fabriksysteme.

Levelt, W. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Levelt, W. & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*(1), 78-106.

Levow, G.-A. (1998). Characterizing and recognizing spoken corrections in human-computer dialogue. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Vol. 1*, (S. 736-742). Abgerufen von http://acl.ldc.upenn.edu/C/C98/C98-1117.pdf

Lewis, J. & Sauro, J. (2009). The factor structure of the system usability scale. In *Proceedings of the Human Computer Interaction International Conference,* (S. 94-103). Abgerufen von gate.ac.uk/sale/dd/statistics/Lewis_Sauro_HCII2009_SUS.pdf

Libermann , A., Cooper, F., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-61.

Linell, P. (1998). *Approaching Dialogue: Talk, Interaction, and Contexts in a Dialogical Perspective.* Amsterdam: John Benjamins Publishing Company.

Litman, D. J. & Pan, S. (2002). Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction, 12*, 111-137.

Litman, D. & Pan, S. (1999). Empirically Evaluating an Adaptive Spoken Dialogue System. In *Proceedings of the 7th International Conference on user Modeling (UM´99),* (S. 55-64). Abgerufen von http://arxiv.org/pdf/cs/9903008.pdf

Maciej, J. & Vollrath, M. (2009). Comparison of manual vs. speech-basedinteraction with in-vehicle information systems. *Accident Analysis and Prevention*, *41*(5), 924-930.

Mann, S. (2010). *User Concepts for In-Car Speech Dialogue Systems and their Integration into a Multimodal Human-Machine Interface* (Dissertation). Universität Stuttgart, Deutschland.

Martens, M. H. & van Winsum, W. (2002). *Measuring distraction: The Peripheral Detection Task*. Abgerufen von http://www-nrd.nhtsa.dot.gov/departments/Human%20Factors/driver-distraction/PDF/34.PDF

Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. In H. Strasser, K. Kluth , H. Rausch & H. Bubb (Hrsg.), *Quality of work and products in enterprises of the future* (S. 57-60). Stuttgart: Ergonomia.

Maurer, R. & Tindall, J. (1983). Effect of postural congruence on client´s perception of counselor empathy. *Journal of Counseling Psychology, 30*(2), 158-163.

McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

McTear, M. (2002). *Spoken dialogue technology. Enabling the conversationaluser interface.* Abgerufen von http://www.ling.helsinki.fi/kit/2002s/ctl190net/materiaali/p90-mctear.pdf

McTear, M. (2004). *Spoken dialogue technology: Toward the conversational user interface.* London: Springer.

Merat, N. (2003). Loading Drivers to Their Limit: The Effect of Increasing Secondary Task on Driving. In *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, (S. 13-18). Abgerufen von http://drivingassessment.uiowa.edu/DA2003/pdf/3_Meratformat.pdf

Metzing, C. & Brennan, S. (2003). When conceptual pacts are broken: Partner-specific effects in the combrehension of referring expressions. *Journal of Memory and Language, 49*(2), 201-213.

Meyer, D. & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review, 63*, 81-97.

Möller, J.-U. (1999). *Dia-MoLE: Modellierung gesprochen-sprachlicher Dialoge unter Zuhilfenahme eines maschinellen Lernverfahrens.* Dissertation, Universität Hamburg: Infix.

Moray, N. (1999). Commentary on Goodman, Tijerina, Bents, and Wierwille, “Using cellular telephones in vehicles: Safe or unsafe?”. *Transportation Human Factors*, *1*(1), 43-46.

Morgan, D., Balentine, B. & William, S. (2001). *How to Build a Speech Recognition Application: Second Edition: A Style Guide for Telephony Dialogues* (2. Ausg.). California: Enterprise Integration Group.

MTV Networks Germany GmbH, Volkswagen AG. (2010). *MePublic – A Global Study on Social Media Youth.* Abgerufen von www.beviacom.de/media/6_research/studien_pdfs/MePublic+Komplett+(Deutsch+pdf)

Mukamel, R., Ekstrom, A., Kaplan, J., Lacoboni, M. & Fried, I. (2010). Single-Neuron Responses in Human during Execution and Observation of Actions. *Current Biology, 20*(8), 750-756.

Nass, C. & Brave, S. (2005). *Wired for speech. How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press.

Nenkova, A., Gravano, A. & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers* (S. 169-172). Abgerufen von www1.cs.columbia.edu/nlp/papers/2008/nenkova_al_08.pdf

Nigay, L. & Coutaz, J. (1993). A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In *Proceedings of INTERCHI '93,* (S. 172-178). Abgerufen von http://www.google.de/url?sa=t&rct=j&q=a%20design%20space%20for%20multimodal%20systems%20concurrent%20processing%20and%20data%20fusion&source=web&cd=1&cad=rja&ved=0CCkQFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.18.150%26rep%3Drep1%26type%3Dpdf&ei=BNSDUK25B8_PsgbP2ICAAw&usg=AFQjCNF-Sz4jwZ3F45EiONh2bdc5vr3yuA

Nordholm, S., Claesson, I. & Grbic, N. (2001). Adaptive Microphone Arrays for Speech Input in Automobiles. In M. Brandstein & D. Ward (Hrsg.), *Microphone Arrays* (S. 307-329). Heidelberg: Springer.

Noszko, T. & Zimmer, A. (2002). Dialog Design für sprachliche und multimodale Mensch-Maschine-Interaktionen im Automobil. In Proceedings of the *38. BDP-Kongress für Verkehrspsychologie Universität Regensburg 2002 Arbeitsgruppe 7: Verkehr und Information: Die Optimierung des Mobilitätssystems,* (S. 1-12). Abgerufen von http://psydok.sulb.uni-saarland.de/volltexte/2006/722/pdf/nozsko.pdf

Okato, Y., Kato, K., Yamamoto, M. & Itahashi, S. (1998). System-User Interaction and Response Strategy in Spoken Dialogue System. In *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, (S. 495-498). Abgerufen von http://www.mirlab.org/conference_papers/International_Conference/ICSLP%201998/PDF/AUTHOR/SL980683.PDF

Opfer, H. & Seitz, G. (2007). *Richtlautsprecher zur Dialogunterstützung im Fahrzeug* Konferenzbeitrag Deutsche Jahrestagung für Akustik (DAGA). Berlin: DEGA.

Oviatt, S., Darves, C. & Coulston, R. (2004). Toward adaptive conversational interfaces: modelling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction, 11*(3), 300-328.

Owens, J., McLaughlin, S. & Sudweeks, J. (2010). On-Road Comparison of Driving Performance Measures When Using Handheld and Voice-Control Interfaces for Mobile Phones and Portable Music Players. *SAE Int. J. Passeng. Cars - Mech. Syst., 1*, 734-743.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Accoustical Society of America*, *119*(4), 2382-2393.

Pauzié, A. (2008). Evaluating driver mental workload using the driving activity load index (DALI). In *Proceedings of the European Conference on Human Interface Design for Intelligent Transport Systems,* (S. 67-77). Abgerufen von http://www.conference.noehumanist.org/articles/Proceedings-HUMANIST-S2.2.pdf

Pauzié, A. & Pachiaudi, G. (1997). Subjective evaluation of the mental workload in the driving context. In T. Rothengatter & V. Carbonell (Hrsg.), *Traffic & Transport Psychology: Theory and Application* (S. 173-182). Oxford: Pergamon.

Pauzié, A., Manzano, J. & Dapzol, N. (2007). Drivers's Behavior and Workload Assessment for New In-Vehicle Technologies Design. In *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, (S. 572-580). Abgerufen von http://drivingassessment.uiowa.edu/DA2007/PDF/092_Pauzie.pdf

Pearson, J., Hu, J., Branigan, H., Pickering, M. & Nass, C. (2006). Adaptive Language behavior in HCI: how expectations and beliefs about a system affect user´s word choice.In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 1177-1180). New York: ACM.

Peissner, M. (2002). What the Relationship between Correct Recognition Rates and Usability Measures Can Tell Us About the Quality of a Speech Application. In *Proceedings of the 6th International Scientific Conference on Work With Display Units,* (S. 296-298). Abgerufen von http://www.google.de/url?sa=t&rct=j&q=what%20realtionship%20between%20correct%20recognition&source=web&cd=2&ved=0CC8QFjAB&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.58.1450%26rep%3Drep1%26type%3Dpdf&ei=2daDUJyvB8HetAbulIDICg&usg=AFQjCNEV9k9KT2U98Rk9khloHIvAgcTvRQ&cad=rja

Peissner, M., Biesterfeld, J. & Heidmann, F. (2004). *Akzeptanz und Usability von Sprachapplikationen in Deutschland.* Stuttgart: Fraunhofer IRB Verlag.

Picheny, M. A., Durlach, N. I. & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28*(1), 96-103.

Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-190.

Pickering, M. & Branigan, H. (1998). The representation of verbs: evidence from syntactic priming in language production. *Journal of Memory and Language, 39*(4), 633-651.

Pickering, M. & Ferreira, V. (2008). Structural priming: a ciritcal review. *Psychological Bullentin, 134*(3), 427-459.

Porzel, R. (2006). How Computers (Should) Talk to Humans. In *Proceedings of Workshop on "How People Talk to Computers, Robots, and Other Artificial Communication Partners"* (S. 7-38). Abgerufen von http://www.sfbtr8.spatial-cognition.de/papers/SFB_TR_8_Rep_010-09_2006.pdf

Porzel, R., Scheffler, A. & Malaka, R. (2006). How entrainment increases dialogical effectiveness. In *Proceedings of the IUI'06 Workshop on Effective Multimodal Dialogue Interaction* (S. 1-8). Abgerufen von http://ww2.cs.mu.oz.au/~lcavedon/finalpdfs/porzel.pdf

Ranney, T. A. (2008). *Driver distraction: a review of the current state-of-knowledge.* Washington, D.C.: U.S. Dept. of Transportation, National Highway Traffic Safety Administration.

Reitter, D. & Moore , J. (2007). Predicting Success in Dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (S. 808-815). Prague: Association for Computational Linguistics.

Richards, M. & Underwood, K. (1984). Talking to Machines: How are People naturally inclined to speak? *Contemporary Ergonomics 1984: Proceedings of the Ergonomics Society's Conference* (S. 62-67). London: Taylor & Francis.

Rizzolatti, G. & Craighero , L. (2004). The mirror-neuron system. *Annual Review Neuroscience. 27*, 169-92.

Röder, V. & Wank, A.-K. (2011). *Wie verhalten sich Autofahrer unter zusätzlicher Belastung? Fahrzeugführung und Reaktionsverhalten in Abhängigkeit von kognitiv unterschiedlich beanspruchenden Zweitaufgaben.* Bachelorarbeit, Chemnitz: TU Chemnitz.

Sacks, H., Schegloff, E. & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language, 50*(4), 696-735.

Saffer, D. (2007). *Designing for Interaction. Creating Smart Applications and Clever Devices.* USA: New Riders Publishing.

Salmen, A. (2002). *Multimodale Menüausgabe im Fahrzeug.* Regensburg: Herbert Utz Verlag.

Sauro, J. (2011). *Measuring Usability with the System Usability Scale (SUS)*. von http://www.measuringusability.com/sus.php

Schenk, J. & Rigoll, G. (2010). *Mensch-Maschine-Kommunikation.* Berlin: Springer.

Schiel, F., Draxler, C. & Libo, M. (2006). Lingua machinae - an unorthodox proposal. In *Proceedings of the INTERSPEECH.* (S. 1-4) Abgerufen von http://www.phonetik.uni-muenchen.de/forschung/publikationen/Schiel-IS2006.pdf

Schmandt, C. (1994). *Voice Communication with Computers.* New York: Van Nostrand Reinhold.

Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition, 47*(1), 1-24.

Schreiner, C. (2011). *Widespread use of SMS and Email while driving.* Abgerufen von http://www.strategyanalytics.com/default.aspx?mod=reportabstractviewer&a0=6149

Selcon, S. & Taylor, R. (1995). Integrating multiple information sources: Using redundancy in the design of warnings. *Ergonomics, 38 (11)*, 2362-2370.

Shannon, C. E. & Weaver, W. (1976). *Mathematische Grundlagen der Informationstheorie.* München: R. Oldenbourg Verlag.

Shneidermann, B. (2002). *User Interface Design.* Bonn: mitp Verlag.

Shriberg, E., Wade, E. & Price, P. (1991). Human–machine problem solving using spoken language systems (SLS): factors affecting performance and user satisfaction. In *Proceedings of the workshop Speech and Natural Language* (S. 49-54). Harriman, NY: Association for Computational Linguistics.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Smith, R. & Hipp, D. (1994). *Spoken Natural Language Dialog Systems: A Practical Approach.* New York: Oxford University Press.

Smith, V. L. & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32,* 25–38.

Sodnik, J., Dicke, C., Tomazic, S. & Billinghurst, M. A. (2008). A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies, 66*, 318-332.

Soltau, H. & Waibel, A. (1998). On the influence of hyperarticulated speech on recognition performance. *Proc. Internat. Conf. Spoken Language Process* (S. 229-232). Beijing, China: International Speech Communication Association.

Stalnaker, R. (1978). Assertion. In P. Cole, *Pragmatics (Syntax & Semtantics 9)* (S. 315-332). New York: Academic Press.

Stent, A., Huffman, M. K. & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50*(3), 163–178.

Suzuki, N. & Katagiri, Y. (2007). Prosodic alignment in human computer-interaction. *Connection Science, 19*(2), 131-141.

Tannen, D. (1987). Repetition in conversation: toward a poetics of talk. *Language, 63*(3), 574-605.

Tomko, S. & Rosenfeld, R. (2006). Shaping user input in speech graffiti: a first pass. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems* (S. 1439-1444). New York: ACM.

Totzke, I. (2001). Die Humanisierung multimodaler HMI - die Bedeutung der Modalität und Situation der Mensch-Maschine-Interaktion. In *Proceedings of the 2. IIR-Fachkongress E-Car Infotainmentplattform & Telematikdienste für das Multimedia Auto,* (S. 1-8). Abgerufen von http://www.psychologie.uni-wuerzburg.de/methoden/texte/2001_totzke_Die_Humanisierung_multimodaler_HMI.pdf

Traum, D. R. & Heeman, P. A. (1996). Utterance Units and Grounding in Spoken Dialogue. In *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, (S. 1884-1887). Abgerufen von http://www.asel.udel.edu/icslp/cdrom/vol3/731/a731.pdf

Tsimhoni, O., Smith, D. & Green, P. (2004). Address entry whiledriving: Speech recognition versus touch-screen keyboard. *Human Factors, 46*(4), 600–610.

Tsimhoni, O., Green, P. & Lai, J. (2001). Listening to natural and synthesized speech while driving: effects on user performance. *International Journal of Speech Technology, 4*(2), 155–169.

Tucholsky, K. (1961). *Gesammelte Werke. Bd. 2.* Reinbek.

Ulich, E. (2005). *Arbeitspsychologie.* Stuttgart: Schäffer Poeschel.

van Baaren, R., Holland, R., Steenaert, B. & van Knippenberg, A. (2003). Mimicry for money: behavioral consequences of imitation. *Journal of Experimental Social Psychology, 39*(4), 393-398.

Vital, F. (2009). Car drivers preferences & expectations towards speech interfaces. In *Proceedings of Nuance Automotive Usability Forum 2009.* St. Goar.

Vollrath, M. (2007). Speech and driving-solution or problem? *Intelligent Transport Systems, IET, 1*(2*),* 89–94.

Vollrath, M. & Krems, J. (2011). *Verkehrspsychologie. Ein Lehrbuch für Psychologen, Ingenieure und Informatiker.* Stuttgart: Kohlhammer Standards Psychologie.

Vollrath, M. & Totzke, I. (2000). In-vehicle communication and driving: an attempt to overcome their interference. *Driver Distraction Internet Forum sponsored by the United States Department of Transportation*. Abgerufen von http://www.psychologie.uni-wuerzburg.de/methoden/texte/2000_vollrath_totzke_In_vehicle_communication_and_driving.pdf

Wade, E., Shriberg, E. & Price, P. (1992). User behaviors affecting speech recognition. In *Proceedings of the ICSLP 92* (S. 995-998). Banff.

Wahlster, W. (2000). *Sprachtechnologie im Alltag: Der Computer als Dialogpartner.* Abgerufen von http://www.dfki.de/~wahlster/Alltag/Alltag.html

Ward, K. & Heeman, P. A. (2000). Acknowledgments in Human-Computer Interaction. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, (S. 280-287). Abgerufen von http://aclweb.org/anthology-new/A/A00/A00-2037.pdf

Ward, N. (1996). Using Prosodic Cues to Decide when to Produce Back-Channel Utterances. In *Proceedings of the ICSLP 96*, (S. 1728-1731). Abgerufen von http://reference.kfupm.edu.sa/content/u/s/using_prosodic_clues_to_decide_when_to_p_96075.pdf

Ward, N. & Nakagawa, S. (2002). Automatic user-adaptive speaking rate selection for information delivery. In *Proceedings of the INTERSPEECH*, (S. 549-552). Abgerufen von http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.35&rep=rep1&type=pdf

Watanuki, K., Sakamoto, K. & Togawa, F. (1994). Analysis of Multimodal Interaction Data in Human Communication. *ICSLP94*, (S. 899-902).

Wechsung, I., Schaffer, S., Schleicher, R., Naumann, A. & Möller, S. (2010). The Influence of Expertise and Efficiency on Modality Selection Strategies and Perceived Mental Effort. In *Proceedings of the INTERSPEECH* (S. 1930-1933). Japan: Curran Associates.

Weizenbaum, J. (1966). ELIZA - A Computer Programm For the Study of Natural Language Communication Between Man and Machine. *Communication of the ACM, 9*(1), 36-45.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159–177.

Wierwille, W. & Tijerina, L. (1995). Eine Analyse von Unfallberichten als ein Mittel zur Bestimmung von Problemen, die durch die Verteilung der visuellen Aufmerksamkeit und der visuellen Belastung innerhalb des Fahrzeugs verursacht werden. *Zeitschrift für Verkehrssicherheit, 41*(4), 164-168.

Wierwille, W., Hulse, M., Fischer, T. & Dingus, T. (1991). Visual Adaptation of the Driver to High-Demand Driving Situations While Navigating with an In-Car Navigation System. In A. Gale, I. Brown, C. Haslegrave, I. Moorehead & S. Taylor (Hrsg.), *Vision In Vehicles 111* (S. 79-87). Amsterdam: Elsevier.

Wikman, A., Nieminen, T. & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics 41 (3)*, 358-372.

Wilkes-Gibbs, D. (1986). *Collaborative processes of language use in conversation.(Nicht veröffentlichte Dissertation)* Standford University, Vereinigten Staaten von Amerika.

Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität.* Göttingen: Hogrefe.

Wirtz, M. & Nachtigall, C. (2006). *Wahrscheinlichkeitsrechnung und Inferenzstatistik: Statistische Methoden für Psychologen* (Bd. 2). Weinheim: Juventa.

Wolf, A. (2003). *Multivariate Statistik.* Abgerufen von http://www.ukaachen.de/go/show?ID=3963820&DV=0&COMP=download&NAVID=3040530&NAVDV=0

Wrede, B., Buschkaemper, S., Muhl, C. & Rohlfing, K. J. (2006). Analysing Feedback in HRI. In *Proceedings of Workshop on "How People Talk to Computers, Robots, and Other Artificial Communication Partners",* (S. 38-54). Abgerufen von http://www.sfbtr8.spatial-cognition.de/papers/SFB_TR_8_Rep_010-09_2006.pdf

Yello Strom. (01. Februar 2012). *Yello Strom Reklame (Markt) [Videodatei]*. Abgerufen von http://www.youtube.com/watch?v=P5mJQn4JfNI

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society* (S. 567-578). Chicago: Chicago Linguistic Society.

Young, K., Regan, M. & Hammer, M. (2003). *Driver distraction: A review of the literature.* Abgerufen von MONASH University Accident Research Centre: http://www.monash.edu.au/miri/research/reports/muarc206.pdf

Zimbardo, P., Gerrig, R., Hoppe-Graf, S. & Engel, I. (1999). *Psychologie.* Berlin: Springer Verlag.

Zoeppritz, M. (1985). *Computer talk?* (Technical Report 85.05). Heidelberg: IBM Scientific Center.

Zoltan-Ford, E. (1991). How to get people to say and typewhat computers can understand. *International Journal of Man-Machine Studies 34*(4), 527–547.

Zwahlen, H. T., Adams, C. C. & DeBald, D. P. (1988). Safety aspects of CRT touch panel controls in automobiles. In A. G. Gale (Hrsg.), *Vision in Vehicles II* (S. 335-344). Amsterdam: Elsevier Science Publishers B.V.

9 Tabellenverzeichnis

10 Abbildungsverzeichnis

11 Abkürzungsverzeichnis

aV	Abhängige Variable
uV	Unabhängige Variable
SDS	Sprachdialogsystem
TTE	Talking too early (Fehlbedienung, zu frühe Spracheingabe)
CG	Common Ground
GC	Grounding Criterion
SGC	System Grounding Criterion
LCE	Least Collaborative Effort
FA	Faktorenanalyse
VL	Versuchsleiter
VP	Versuchsperson
OOV	Out-of-Vocabulary
PTT	Push-to-talk
HMM	Hidden-Markov-Modelle
HMI	Mensch-Maschine-Interaktion
SUS	System Usability Scale
SUMI	Software Usability Measurement Inventory
DALI	Driving Activity Load Index
LCT	Lane Change Task
TTS	Text-to-Speech
DALI	Driving Activity Load Index

12 Anhang

Zusammenfassung

Diese Arbeit adressiert die Implementation zwischenmenschlicher Dialogprinzipien im Rahmen der Gestaltung automotiver Sprachdialogsystemen (SDS). Der Transfer der kollaborativen Strategien, insbesondere die kontinuierliche, nutzer- und situationsabhängige Vermittlung von Feedback soll Gegenstand von empirischen Untersuchungen sein.

Bisher kann eine deutliche Diskrepanz zwischen Kundenerwartungen und technischer Realität festgestellt werden. Obwohl in den letzten Jahrzehnten deutliche Verbesserungen der Spracherkennungstechnologie erreicht werden konnten, übernehmen aktuelle SDS die kooperative Verantwortung des Empfängers, dem Sprecher Indizien über die eigenen Verstehensprozesse zu präsentieren und den gemeinsamen Aufwand zu minimieren, nur unzureichend. Die vorliegende Dissertation diskutiert nicht-technische Lösungsansätze, die die Anpassung des Systemverhaltens an bestehende Kommunikationsprozesse vorsehen, um die Koordination der Wissensstände zwischen Mensch und Maschine zu ermöglichen.

Drei verschiedene Grounding-Elemente wurden auf die Mensch-Maschine-Interaktion angewendet. Zunächst wurde ein System implementiert, welches visuelle Repräsentationen der Dialoginhalte und -zustände bot. In einer zweiten Umsetzung wurde ein flexibles System Grounding Criterion in Anlehnung an menschliches Rückfrageverhalten umgesetzt, so dass das System nur dann eine Bestätigungsanfrage erbat, wenn es sich unsicher war. Das dritte System adressierte Angleichungsprozesse in dem die Systemausgabe syntaktisch und lexikalisch an die Nutzereingabe angepasst wurde. Um den Einfluss dieser drei Umsetzungen auf Gebrauchstauglichkeitsbeurteilungen zu untersuchen, wurden umfangreiche Nutzerstudien durchgeführt. Während die Versuchspersonen im Fahrsimulator fuhren, nutzten sie die verschiedenen SDS für unterschiedliche Bedienaufgaben aus dem Adressbuchkontext. Die Ergebnisse der empirischen Untersuchungen zeigen, dass die Anpassung von SDS an bestehende Kommunikationsstrategien zu erhöhter Nutzerzufriedenheit trotz technischer Einschränkungen von state-of-the-art Spracherkennungstechnologie führen kann. Die Implementation eines flexiblen Grounding Criterions, das die Effizienz und die Effektivität der Interaktion steigern konnte, stellte dabei den erfolgreichsten Transfer von zwischenmenschlichen Dialogstrategien auf den Mensch-Maschine-Dialog dar.

Schlagwörter: Mensch-Maschine-Interaktion, Sprachdialogsystem, Grounding, Usability

Abstract

This work addresses the evaluation of speech dialog systems (SDS) that make use of collaborative strategies from human dialog by providing continuous and appropriate feedback whilst showing adaptive interaction structures.

Users' experience with today's spoken dialog systems is characterized by interaction structures which do not meet their expectations. The fact that users feel uncomfortable while interacting with current systems can be explained as failed grounding processes, in which users lack evidence to coordinate their knowledge states with the SDS. This thesis proposes solutions of how to overcome difficulties with in-vehicle speech dialog systems from a non-technical usability point of view by adapting the system behavior to existing communication strategies.

Three different grounding strategies were applied to the human machine dialog. Firstly, a system was implemented that gave visual representation of the dialog content and processes. Secondly, a flexible system grounding criterion was realized, so that the system only asked for confirmation if it was insecure, similar to what humans do. The third implementation was concerned with alignment strategies namely by adapting the system's output syntactically and lexically towards the users' input. User studies were conducted to examine the impact of these three implementations on system usability ratings. While driving the simulator, subjects were using the different SDS for several tasks concerning the address book. The tested SDS differed in visualizations, dialog behavior or prompt design.

The results of the evaluations show, that adapting the SDS to existing communication strategies can lead to improved user satisfaction despite the persisting shortcomings of state-of-the-art speech technology. The implementation of a flexible grounding criterion, which could enhance the efficiency and effectiveness of the interaction, was thereby the most successful transfer from human communication strategies to human machine dialog.

Keywords: Human Machine Interaction, Speech Dialog System, Grounding, Usability